CD-CAT中基于SCAD惩罚和EM视角的在线标定方法开发——基于G-DINA模型*

谭青蓉[1,2]　蔡　艳[2]　汪大勋[2]　罗　芬[3]　涂冬波[2]
(1 陆军军医大学医学心理系基础心理学教研室, 重庆 400000)
(2 江西师范大学心理学院, 南昌 330022)
(3 江西师范大学计算机信息工程学院, 南昌 330022)

摘　要　G-DINA (the generalized deterministic input, noisy and gate)模型限制条件少，应用范围广，满足大量心理与教育评估测验数据的要求。研究提出一种适用于G-DINA等模型的同时标定新题*Q*矩阵与项目参数的认知诊断计算机化自适应测验(CD-CAT)在线标定新方法SCADOCM，以期促进CD-CAT在实践中的推广与应用。本研究分别基于模拟题库以及真实题库进行研究，结果表明：相比传统的SIE方法，SCADOCM在各实验条件下均具有较为理想的标定精度与标定效率，应用前景较好；SIE方法不适用于饱和的G-DINA等模型，其各实验条件下的*Q*矩阵标定精度均较低。

关键词　认知诊断计算机化自适应测验, 在线标定, *Q* 矩阵, G-DINA 模型, SCAD 惩罚
分类号　B841

1 引言

如何高效、准确地为被试提供其在所测内容上详细且有价值的诊断信息以满足被试的测验需求？这是近年来心理与教育测量学研究者和实践者都极为关注的问题。在心理评估中，如果测验能快速、准确、高效地为临床心理医生尤其是新手医生提供来访者在某一心理问题上的具体症状表现，帮助临床医生更好地理解心理问题及一些具体症状之间潜在的复杂关系，心理医生可及时地制定有效的预防和干预策略，推进心理治疗进程(如, de la Torre et al., 2018; Tan et al., 2023)。而在教育测评中，如果测验能快速、准确、高效地为教师提供学生掌握和欠缺的

收稿日期：2022-09-26
* 国家自然科学基金项目(62167004, 32160203, 32300942, 31960186 和 61967009)。
通信作者：蔡艳, E-mail: cy1979123@aliyun.com; 汪大勋, E-mail: wangda.xun@163.com; 涂冬波, E-mail: tudongbo@aliyun.com

具体知识点，教师在课堂上可以重点讲授学生有待提高的知识点，学生也可以针对自己的弱项进行有针对性的学习，从而减轻学生负担，改进教学，提高教学效果(如，Tang & Zhan, 2021)。

认知诊断计算机化自适应测验(cognitive diagnostic computerized adaptive testing, CD-CAT)正是在这一背景下产生，它包含了近来蓬勃发展的认知诊断(cognitive diagnosis, CD)和计算机化自适应测验(computerized adaptive testing, CAT)两种测量技术的优点，是实现以上测量目标较为理想的选择(Cheng, 2009; Lin & Chang, 2019; Xu et al., 2016)。认知诊断的迅速发展，很大程度上取决于实践中对于形成性评估(formative assessment)的需求。不同于仅提供测验总分的总结性评估(summative assessment)，认知诊断为每个被试提供属性掌握模式，该模式详细描述了被试在所测概念或内容上的掌握情况，可为测验后的进一步补救干预提供重要参考(de la Torre, 2011; Junker & Sijtsma, 2001)。CAT 因其量身定制与高效的特点而备受研究者与实践者的青睐。CAT 根据每个被试的潜在特质水平为其定制一个测验，被试作答项目大多都与其潜在特质水平相匹配，因此 CAT 可为被试提供更为有效且精确的潜在特质估计值。而 CD-CAT 同时具备 CAT 的特点以及认知诊断的功能，它通过“量体裁衣”的个性化测验快速准确地探查被试在所测内容上的优势和不足，可及时为被试提供精细的诊断反馈信息，在提高测验结果准确性的同时极大地减轻了测验参与者的作答负担(Chen et al., 2012; Chen et al., 2015; Lin & Chang, 2019; Liu et al., 2013)。这符合“双减”等政策的精神和要求，也较好地满足了当前国家和社会发展的实际需要，有利于促进精准、自适应和个性化的心理与教育测评，以及考试的数字化革新。

CD-CAT 的有效性依赖于高质量的题库(item bank)。然而，在 CD-CAT 持续使用一段时间后，题库中的部分题目会变得过时或者丧失功能，这些题目需及时使用新题予以替换以保证测验和题库的质量(Chen et al., 2012; Chen et al., 2015; Kang et al., 2020)。具体而言，需要邀请经验丰富的领域专家和心理测量学家根据诊断目的编制新题(即待加入题库但未标定参数的题目)，然后估计新题参数，并将其与题库中已有的题目置于同一量尺之上。在线标定(online calibration)技术是 CAT 中一种有效的项目增补方法，它是指在测验过程中，让被试同时作答新题与旧题(题库中已有的已标定参数的题目)，并根据其作答来标定新题参数的过程(陈

平，辛涛, 2011a)。除可节约资源投入且相同测量模式使得被试作答新题和旧题的动机相同这些优势外，在线标定的另一重要优势是无需复杂的等值技术以用于解决大型题库构建时所面临的测验等值等具有挑战性的难题(Chen & Wang, 2015; Chen et al., 2012)。至今为止，在单维计算机化自适应测验(unidimensional CAT, UCAT)以及多维计算机化自适应测验(multidimensional CAT, MCAT)领域中，研究者已提出了多种高效的在线标定方法。如，方法 A (Method A; Stocking, 1988)、一个 EM 循环的边际极大似然估计方法(marginal maximum likelihood estimate with one EM cycle, OEM; Wainer & Mislevy, 1990)、多个 EM 循环的边际极大似然估计方法(marginal maximum likelihood estimate with multiple EM cycles, MEM; Ban et al., 2001)、FFMLE-Method A 方法(陈平, 2016)、M-Method A 方法(Chen et al., 2017)、M-MEM-BME 方法(Chen, 2017)等。

CD-CAT 中可使用在线标定技术标定新题的参数，但有一个问题值得思考，即认知诊断测验中是否需要进行等值，是否有必要使用在线标定技术对新题进行标定？de la Torre 和 Lee (2010)在研究中指出当模型与数据完全拟合时，决定型输入噪音与门(the deterministic input, noisy and gate, DINA; Junker & Sijtsma, 2001)模型的项目参数具有不变性；Bradshaw 和 Madsion (2015)，Madsion 和 Bradshaw (2018)也在其研究中指出对数线性认知诊断模型(log-linear cognitive diagnosis model CDM, LCDM; Henson et al., 2009)和基于 LCDM 模型开发的 TDCM (the Transition Diagnostic Classification Model)在模型与数据拟合的情况下参数具有不变性。在此条件下，无需通过等值来保证被试参数估计值在同一量尺上。然而，其研究也指出在模型与数据不完全拟合时，难以观察到参数不变性；且即使模型与数据拟合的情况下，参数不变性也会随着标定样本的减少而减弱(Bradshaw & Madsion, 2015; de la Torre & Lee, 2010; Madsion & Bradshaw, 2018)。这表明参数不变性成立需满足一些必备的条件：如模型与数据完全拟合，标定样本量足够大(如不少于 1000)，在这些条件下可以不进行等值。但在实际测验情境中，模型与数据完全拟合的情况并不总能得到满足，且在同一次测验中也较难获得足够大的标定样本，这都会导致项目参数估计出现偏差，影响被试的分类准确性和 Q 矩阵的标定正确性。因此，在 CD-CAT 题库建设中有必要进行在线标定，这有利于降低项目参数估计偏差等所带来的影响，提高 CD-CAT 题库和测验的质量。

目前，CD-CAT 中有关在线标定方法的研究仍然较为薄弱，而且不同于 UCAT 和 MCAT，CD-CAT 中标定新题时不仅需要考虑新题项目参数的标定，还需考虑新题 *Q* 矩阵的标定。*Q* 矩阵作为认知诊断的核心成分，在大多数情况下是未知的。在实际测验中，*Q* 矩阵一般由领域专家和心理测量学专家共同界定，需要耗费大量的人力和物力资源。另外，由专家界定的 *Q* 矩阵容易受专家主观因素的影响造成错误界定，而 *Q* 矩阵的错误界定最终影响项目参数估计精度和被试分类准确性(de la Torre & Chiu, 2016; Rupp & Templin, 2008)。因此，新题 *Q* 矩阵的标定是 CD-CAT 中标定新题时不容忽视的一个方面。

截至目前，已有部分研究对 CD-CAT 中新题 *Q* 矩阵与项目参数的同时标定进行了探索。例如，陈平和辛涛(2011b)提出的联合估计算法(joint estimation algorithm, JEA)，Chen 等人(2015)提出的 SIE (single-item estimation)方法,谭青蓉等人(2021)提出的基于熵的信息增益在线标定方法(Information Gain of Entropy-based Online Calibration Method, IGEOCM)，以及 Tan 等人(2022)提出的基于基尼的方法(the Gini-based method)等均为同时标定新题 *Q* 矩阵与项目参数的在线标定方法。已有研究表明 JEA、SIE、IGEOCM 和基于基尼的方法等在 DINA 模型下具有较为理想的项目标定精度，但在其它模型尤其是适用面更广、限制条件非常少的饱和认知诊断模型(如拓广的 DINA 模型，即 G-DINA; de la Torre, 2011)下的性能仍有待进一步考察。

相比于 DINA 模型，G-DINA 等模型因限制条件少而有着更广的适用范围，能满足心理与教育评估中多数测验数据的要求(de la Torre, 2011; de la Torre et al., 2018; Tu et al., 2017; Xi et al., 2020)，在实践研究中的应用日益广泛。如心理临床诊断评估中，只要被试符合心理障碍诊断标准中的部分症状便可实现对被试的临床诊断。以网络成瘾为例，《精神障碍诊断与统计手册》第五版(the 5th edition of the diagnostic and statistical manual of mental disorders, DSM-V)中界定了网络成瘾的 9 条症状标准，被试符合其中 5 条及 5 条以上症状可诊断为网络成瘾。此时，DINA 模型显然不适用于此类测验，它假定被试在项目上的作答只受到项目测量的所有属性的交互作用影响，而不受主效应及其它类型的交互作用的影响。如果强行使用该模型来分析整个测验可能导致数据与所用模型的不适配，继而影响诊断结果的可信性和精确性(Hou, 2013)。而 G-DINA 模型则没有这些严格的假设，

认为被试的作答可以是由项目测量的各属性的主效应与各种类型的交互效应的共同影响，如果主效应(或交互效应)的系数估计值为 0 或接近 0，则此时主效应(或交互效应)的作用不明显，即此时不存在主效应(或交互效应)，但若系数显著不为 0，则说明存在主效应(或交互效应)，因此 G-DINA 模型更为灵活，更适合该类测验。

然而，及至目前尚未有公开发表的期刊文章研究应用于限制条件少的 G-DINA 等模型的 Q 矩阵与项目参数同时性在线标定方法，这在一定程度上限制了 CD-CAT 在实践中的应用范围，阻碍了 CD-CAT 在实际测验中的进一步推广。鉴于此，研究拟引入数据挖掘中 SCAD (smoothly clipped absolute deviation penalty, SCAD; Fan & Li, 2001)方法选择特征的思路提出一种适用于 G-DINA 等模型的 Q 矩阵与项目参数同时性在线标定方法，旨在为 CD-CAT 在实践中的进一步推广与应用提供高效准确的方法学支持。

2 G-DINA模型及SIE方法简介

2.1 G-DINA 模型

已有认知诊断模型中，基于 DINA 模型拓展而来的 G-DINA 模型是一个限制条件少，应用范围更广的模型，符合大量心理与教育评估测验数据的要求，在实践中所受到的重视日益增加，越来越多的研究者基于 G-DINA 模型开发认知诊断测验(如, de la Torre et al., 2018; Tu et al., 2017; Xi et al., 2020)。故研究在 G-DINA 模型框架下介绍新的在线标定方法并对其进行验证，该新方法同样可以应用于其它认知诊断模型。

令测验测量的属性个数为 K，$\boldsymbol{q}_j=(q_{j1},\cdots,q_{jK})$ 为项目 j 的 q 向量，是测验 Q 矩阵的第 j 行，若被试正确作答项目 j 需要掌握第 k 个属性，$q_{jk}=1$，否则 $q_{jk}=0$；X_{ij} 表示被试 i 在项目 j 上的作答；$\boldsymbol{\alpha}_c=(\alpha_{c1},\alpha_{c2},\cdots,\alpha_{cK})$ 表示第 c 类属性掌握模式，其中 α_{ck} 表示第 c 类属性掌握模式下的被试是否掌握第 k 个属性，若掌握了第 k 个属性，$\alpha_{ck}=1$，否则 $\alpha_{ck}=0$。G-DINA 模型认为属性掌握模式不同的被试在项

目上的正确作答概率并不一致，将被试分为$2^{K_j^*}$个类别，其中$K_j^* = \sum_{k=1}^{K} q_{jk}$表示项目$j$测量的属性个数。根据所用链接函数的不同，G-DINA 模型有不同的数学表达式，其中最为常用的链接函数为对数链接函数(log link function)、logit 链接函数(logit link function)和一致性链接函数(identity link function)。而一致性连接函数下的 G-DINA 模型，是 G-DINA 模型更为一般化的形式(de la Torre, 2011)，其数学表达式可写为：

$$P(\boldsymbol{\alpha}_{cj}^*) = P(\boldsymbol{X}_j = 1|\boldsymbol{\alpha}_{cj}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk}\alpha_{ck} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'}\alpha_{ck}\alpha_{ck'} + \cdots + \delta_{j12\cdots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{ck}, \quad (1)$$

上式中，$\boldsymbol{\alpha}_{cj}^*$表示基于项目 j 所测量属性的缩减属性掌握模式，其中$c = 1, 2, \cdots, 2^{K_j^*}$。例如，测验共测量 3 个属性，项目$j$测量了测验的前两个属性$\boldsymbol{q}_j = (1, 1, 0)$，则$K_j^* = 2$，$\boldsymbol{\alpha}_{cj}^* = ((0, 0),(1, 0),(0, 1),(1, 1))^T$；$\delta_{j0}$表示项目$j$的截距参数，也称之为基线概率，指被试未掌握项目$j$测量的所有属性但在该项目上作答正确的概率，其为非负值；δ_{jk}表示项目j上属性k的主效应，指被试掌握属性k对被试正确作答该项目概率的增加效应，一般取非负值，值越大说明掌握该属性对于正确作答该项目的贡献越大；$\delta_{jkk'}$表示项目j上属性k和k'的交互效应，$\delta_{j12\cdots K_j^*}$是所有属性的交互效应。文中使用$\boldsymbol{\delta}_j$表示项目j的项目参数向量，G-DINA 模型中$\boldsymbol{\delta}_j = (\delta_{j0}, \delta_{j1}, \cdots, \delta_{jK_j^*}, \delta_{j12}, \cdots, \delta_{j(K_j^*-1)(K_j^*)}, \cdots, \delta_{j12\cdots K_j^*})$。

2.2 SIE 方法

CD-CAT中已有的同时标定新题Q矩阵与项目参数的方法主要包含了JEA (陈平, 辛涛, 2011b)、SIE (Chen et al., 2015)、IGEOCM (谭青蓉 等, 2021)和基于基尼的方法(Tan et al., 2022)等。其中，JEA方法在项目质量高且样本量大时具有较高的项目标定精度，但其在项目质量较低时的项目标定精度仍有待于进一步提高。而实际测验题库中，可能既包含了质量高的项目，也包含了质量低的项目。如Liu等人(2013)开发的中国大型英语二级测验题库，其项目失误参数(被试掌握了项目测量的所有属性但错误作答该项目的概率)的范围在0.001到0.5之间。在新题

的质量较低时，若使用JEA方法来标定新题，可能导致新题的标定精度较低，从而影响整个题库以及测验的质量。另外，理论上IGEOCM和基于基尼的方法可用于DINA模型外的其它认知诊断模型，但该类方法受被试类别数量的影响，DINA模型在每个项目上均将被试区分为两个类别，而G-DINA模型在每个项目上将被试区分为$2^{K_j^*}$(K_j^*表示项目j测量的属性个数)个类别，其在G-DINA等模型下的性能可能并不理想。如G-DINA等模型下，被试类别随项目测量属性个数的增加而增加，而熵的信息增益指标会随着被试类别的增加而增加(李航, 2012)。因此，在G-DINA等模型下使用IGEOCM方法标定新题 q 向量，可能出现属性指定过多的情况。基于以上分析，文中仅详细介绍SIE方法，并将其与新方法进行比较。

SIE 方法基于 DINA 模型提出，其在标定新题时考虑了被试属性掌握模式的估计误差，标定新题 Q 矩阵和项目参数时充分利用被试的属性掌握模式后验分布(Chen et al., 2015)。SIE 方法标定新题时包含了 Q 矩阵标定和项目参数标定两个部分。对于新题 Q 矩阵的标定，首先基于被试在旧题上的作答计算作答了新题 j 的被试的属性掌握模式后验分布。随后，根据被试属性掌握模式后验分布及每种属性掌握模式在 q 向量为$\boldsymbol{q}_j$的新题 j 上的正确作答概率计算具有某一特定作答R_{ij}的被试 i 的后验预测分布：

$$P_i(\boldsymbol{q}_j,\boldsymbol{\delta}_j)=P(R_{ij}=1|\boldsymbol{q}_j,\boldsymbol{\delta}_j)=\sum_{c=1}^{2^K}\pi_{ij}(\boldsymbol{\alpha}_c)P(\boldsymbol{q}_j,\boldsymbol{\delta}_j|\boldsymbol{\alpha}_c), \tag{2}$$

其中$\boldsymbol{\delta}_j$表示项目参数向量，DINA 模型下包含失误参数s_j和猜测参数g_j；$P(\boldsymbol{q}_j,\boldsymbol{\delta}_j|\boldsymbol{\alpha}_c)$表示属性掌握模式为$\boldsymbol{\alpha}_c$的被试在新题 j 上的正确作答概率；$\pi_{ij}(\boldsymbol{\alpha}_c)$表示作答了新题 j 的被试 i 的属性掌握模式为$\boldsymbol{\alpha}_c$的后验概率，基于被试 i 在 O 个旧题上的作答($\boldsymbol{U}_i, i=1, 2, \cdots, n_j$)计算获得：

$$\pi_{ij}(\boldsymbol{\alpha}_c)=\frac{\pi(\boldsymbol{\alpha}_c)\prod_{o=1}^{O}P(\boldsymbol{q}_o,\boldsymbol{\delta}_o|\boldsymbol{\alpha}_c)^{U_{io}}[1-P(\boldsymbol{q}_o,\boldsymbol{\delta}_o|\boldsymbol{\alpha}_c)]^{1-U_{io}}}{\sum_{c=1}^{2^K}\pi(\boldsymbol{\alpha}_c)\prod_{o=1}^{O}P(\boldsymbol{q}_o,\boldsymbol{\delta}_o|\boldsymbol{\alpha}_c)^{U_{io}}[1-P(\boldsymbol{q}_o,\boldsymbol{\delta}_o|\boldsymbol{\alpha}_c)]^{1-U_{io}}}, \tag{3}$$

上式中，$\pi(\boldsymbol{\alpha}_c)$表示属性掌握模式为$\boldsymbol{\alpha}_c$的先验概率，$P(\boldsymbol{q}_o,\boldsymbol{\delta}_o|\boldsymbol{\alpha}_c)$表示属性掌握模

式为$\boldsymbol{\alpha}_c$的被试在旧题o上的正确作答概率，U_{io}表示被试i在旧题o上的作答。

最后，结合被试后验预测分布及其在新题j上的作答R_{ij}构建似然并最大化似然函数来估计新题的q向量，其表达式如下：

$$\hat{\boldsymbol{q}}_j=\underset{\boldsymbol{q}_j^*\in\boldsymbol{Q}_j}{\operatorname{argmax}}L(\boldsymbol{q}_j^*,\boldsymbol{\delta}_j)=\underset{\boldsymbol{q}_j^*\in\boldsymbol{Q}_j}{\operatorname{argmax}}\left\{\prod_{i=1}^{n_j}P_i(\boldsymbol{q}_j^*,\boldsymbol{\delta}_j)^{R_{ij}}[1-P_i(\boldsymbol{q}_j^*,\boldsymbol{\delta}_j)]^{1-R_{ij}}\right\},\qquad(4)$$

其中，$\boldsymbol{Q}_j$表示新题j所有2^K-1种可能q向量的集合。此外，SIE方法使用EM算法来估计新题的项目参数。

需注意的是，DINA 模型下使用 SIE 方法标定新题时对于任一的项目参数估计值，需将新题的所有可能q向量代入似然函数以计算所有可能q向量所对应的似然值，在此基础上标定新题的q向量与项目参数。这在DINA模型下是可行的，因为该模型下项目参数的个数不随项目所测属性个数的变化而发生变化，不同q向量所对应的项目参数个数均为 2，也即失误参数和猜测参数。但这在 G-DINA 模型下是难以实现的，因为该模型下项目参数的个数随项目所测属性个数的变化而变化，不同q向量所对应的项目参数个数可能不同。如项目测量 2 个属性时，项目参数的个数为 4；而项目测量 3 个属性时，项目参数的个数为 8。因此，将 SIE 方法从 DINA 模型拓展到 G-DINA 模型时，对于根据某一q向量估计的项目参数估计值，仅结合该项目参数估计值及其对应的q向量计算一个似然值。如，基于$\boldsymbol{q}_j=[10010]$估计的项目参数值，仅将其与$\boldsymbol{q}_j=[10010]$结合计算似然值，而不与$\boldsymbol{q}_j=[10011]$等可能的项目q向量结合来计算似然值。对于新题j的所有可能q向量及其各自对应的项目参数估计值，均可以计算一个似然值。若新题的可能q向量个数为 8，则可以计算 8 个似然值，选择最大似然值对应的q向量与项目参数作为新题的q向量与项目参数估计值。除此之外，G-DINA 模型下使用 SIE 方法标定新题时的步骤均与 DINA 模型一致。

3 基于 SCAD 的在线标定方法(SCADOCM)开发

3.1 SCADOCM 开发的基本思想

目前，数据挖掘中多数方法都围绕正则化方法进行，正则化方法是系数收缩方法的一种，通过压缩特征系数来达到特征选择的目的，已成为一种主流的特征选择方法。正则化方法基于惩罚的思想，在目标函数上增加一个惩罚项，使得新目标函数最小化以选择重要特征。SCAD 惩罚是一种正则化方法，其在特征选择上具有良好的性能(Fan & Li, 2001)。为简化表达，将 SCAD 惩罚称为 SCAD，基于 SCAD 的对数似然函数可表示为：

$$SCAD(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - n\sum_{w=1}^{W} p_{\lambda}(|\beta_w|), \tag{5}$$

其中，$l(\boldsymbol{\beta})$表示基于特征构建的回归方程的对数似然函数，若基于特征构建的回归为 logistic 回归，则其对数似然函数可表示为：

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n}[\boldsymbol{R}_i(\boldsymbol{D}_i^T\boldsymbol{\beta}) - \log(1+\exp(\boldsymbol{D}_i^T\boldsymbol{\beta}))], \tag{6}$$

其中 n 表示被试人数，$\boldsymbol{R}_i$ 表示被试 i 在因变量 R 上的作答，$\boldsymbol{D}_i^T$ 表示被试 i 在自变量向量集 D 上的作答向量的转置，$\boldsymbol{\beta}$ 表示回归系数向量。

$n\sum_{w}^{W} p_{\lambda}(|\beta_w|)$为对数似然函数的惩罚项，$W$ 为自变量向量 D 的维数，$p_{\lambda}(.)$为惩罚函数，其形式构造如下：

$$p_{\lambda}(|\boldsymbol{\beta}|) = \begin{cases} \lambda|\boldsymbol{\beta}|, & |\boldsymbol{\beta}| \leqslant \lambda \\ -\dfrac{|\boldsymbol{\beta}|^2 - 2a\lambda|\boldsymbol{\beta}| + \lambda^2}{2(a-1)}, & \lambda \leqslant |\boldsymbol{\beta}| \leqslant a\lambda \\ \dfrac{(a+1)\lambda^2}{2}, & |\boldsymbol{\beta}| > a\lambda \end{cases}, \tag{7}$$

a 和 λ 为 SCAD 函数中需定义数值的两个参数。Fan 和 Li (2001)建议 $a = 2 + \sqrt{3} = 3.7$，该值在各种特征选择问题中都表现出较好的性能。λ是一个调整参数(tuning parameter)，极大程度地影响 SCAD 方法的性能(Fan & Li, 2001; Fan & Lv, 2010; Fan & Tang, 2013; Zhang et al., 2010)。Fan 和 Li (2001)建议$\lambda = 0.7$，研究者也提出了不同的λ参数选择方法，如 GCV 准则、AIC 准则和 BIC 准则等。BIC 准则是较为常用的λ参数选择方法(Wang et al., 2007; Zhang et al., 2010)。

SCAD 对数似然函数第一项表示模型拟合，值越小模型拟合越好；第二项是对模型中所包含的自变量个数(模型复杂度)的惩罚，较好地体现了模型拟合与复杂性的权衡。基于 SCAD 的似然函数可使用局部二次逼近算法(local quadratic approximations, LQA)来估计$\hat{\beta}_\lambda$(Fan & Li, 2001)。LQA 算法的特征在于把收敛于 0 的回归系数估计为 0，从而达到简化模型，提高运算效率的目的。

新题 j 的 q 向量估计可视为一个特征选择问题，将测验测量的所有属性作为待选择的特征，从所有测验属性中选择重要属性作为新题 j 的测验属性，构建 q 向量(q 向量中新题 j 的测验属性标记为 1，其它属性标记为 0)。若项目 j 测量了某几个属性，则在这些属性上掌握概率更高的被试正确作答项目 j 的可能性更大，而在这些属性上掌握概率更低的被试正确作答项目 j 的可能性更小。因此，某一属性的被试掌握概率对被试正确作答的影响越大，说明该属性对于项目来说越重要，反之若某一属性的被试掌握概率对被试正确作答的影响可忽略不计，则说明项目可能未测量该属性。将被试在新题 j 上的作答数据 R 视为因变量，被试在每个测验属性上的掌握情况视为自变量(待选特征)构建 SCAD 对数似然函数，然后最小化该目标函数以选择新题 j 的测验属性，构建新题 q 向量。基于该思路，本研究提出基于 SCAD 的在线标定方法(SCAD-based online calibration method, SCADOCM)，该方法使用 SCAD 方法标定新题的 Q 矩阵，随后使用 EM 算法标定新题的项目参数。SCADOCM 标定新题 Q 矩阵与项目参数的计算公式及其过程详细介绍如下。

3.2 SCADOCM 中 Q 矩阵与项目参数标定的算法设计

本节将详细说明如何使用 SCADOCM 来估计新题的 q 向量与项目参数。对于新题 q 向量的估计，首先将新题的 q 向量估计视为一个特征选择问题，然后通过 SCAD 构造一个有效可行的估计量。在认知诊断中，被试对新题 j 的回答取决于他们对属性的掌握程度。一般来说，掌握新题 j 所测量属性的被试，正确作答新题 j 的概率更高。反之，如果掌握了第 k 个属性的被试在新题 j 上具有更高的正确作答概率，那么新题 j 极有可能测量了属性 k。那么如何才能从测验测量的所有属性中选择显著影响被试正确作答该题的属性呢？SCAD 方法作为一种具有众多优良特性的特征选择方法，是一种可行的解决方案。

基于测验测量属性以及被试在新题上的作答使用 SCAD 方法标定新题 Q 矩阵，首先需构建属性与被试作答间的回归模型。这一步的关键是找到合适的指标来描述考生对属性的掌握程度。被试在测验所测属性上的边际掌握概率可基于 CD-CAT 过程中被试对旧题的作答估计获得，该指标较好地体现了被试对于属性的掌握程度。被试在某个属性上的边际掌握概率越高，则被试掌握该属性的概率越大。此外，被试在新题 j 上的作答服从伯努利分布。因此，对于新题 j，基于被试在所测属性上的边际掌握概率及其在项目上的作答，可构建如下 logistic 回归模型：

$$P(\boldsymbol{R}_j = 1|\boldsymbol{D}) = \frac{\exp(\boldsymbol{D}^T\boldsymbol{\beta})}{1+\exp(\boldsymbol{D}^T\boldsymbol{\beta})}, \tag{8}$$

其中，$\boldsymbol{D}$ 表示大小为 $K \times n_j$ 的被试属性边际掌握概率矩阵，$\boldsymbol{\beta}$ 表示大小为 $K \times 1$ 的属性回归系数向量。

随后，可基于该回归方程构建对数似然函数，其公式可表达如下：

$$l(\boldsymbol{\beta}) = \sum_{i=1}^{n_j}\left[R_{ij}(\boldsymbol{D}_i^T\boldsymbol{\beta}) - \log\left(1+\exp(\boldsymbol{D}_i^T\boldsymbol{\beta})\right)\right], \tag{9}$$

其中 R_{ij} 表示被试 i 在新题 j 上的作答。在公式(9)上增加 SCAD，则可构建基于 SCAD 的对数似然函数如下：

$$SCAD(\boldsymbol{\beta}) = l(\boldsymbol{\beta}) - n_j\sum_{w=1}^{W} p_\lambda(|\beta_w|), \tag{10}$$

其中 $p_\lambda(|\boldsymbol{\beta}|)$ 如公式(7)所示，本研究采用建议的 $a = 2+\sqrt{3} = 3.7$ (Fan & Li, 2001)，使用 BIC 准则选择 λ 参数。对于某一给定 λ 值，BIC 指标可计算如下：

$$BIC(\lambda) = -2l\left(\hat{\boldsymbol{\beta}}_\lambda\right) + |\boldsymbol{v}_\lambda|\log(n_j), \tag{11}$$

其中 $\boldsymbol{v}_\lambda = \left\{k\text{: } \hat{\beta}_{\lambda k} \neq 0\right\}$ 表示不包含截距项的活动集，$|\boldsymbol{v}_\lambda|$ 表示该活动集的大小。

最后，基于 BIC 准则选择的 λ 参数，最小化公式(10)可获得 $\boldsymbol{\beta}$ 的估计值，其表达式为：

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, SCAD(\boldsymbol{\beta}). \tag{12}$$

若 $\hat{\beta}_k \neq 0$，则新题 j 测量了属性 k。例如，K=5，$\hat{\boldsymbol{\beta}}$ 中的第一个元素和第四个元素

为非 0 系数，则新题 j 的 q 向量为$\boldsymbol{q}_j = [10010]$。若对于 BIC 准则所选择的$\lambda$参数出现$\hat{\boldsymbol{\beta}} = 0$，则选择$\lambda$参数取最小值时所获得的回归系数估计值中最大回归系数所对应的属性作为新题 j 的测验属性以确保新题 j 至少测量一个属性。λ参数的取值区间参考 Breheny 和 Huang (2011)研究所提方法。

SCADOCM 中，在使用 SCAD 方法标定新题的 q 向量之后，需要根据该 q 向量来估计其项目参数，具体为采用 EM 算法来估计新题的项目参数(Chen et al., 2015)。在 E 步中, 首先基于被试 i 在新题 j 上的作答 R_{ij} 计算每个被试的后验分布，其公式如下：

$$Post_{ij}(\boldsymbol{\alpha}_c) = \frac{\pi_{ij}(\boldsymbol{\alpha}_c) P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)^{R_{ij}} [1 - P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)]^{1-R_{ij}}}{\sum_{c=1}^{2^K} \pi_{ij}(\boldsymbol{\alpha}_c) P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)^{R_{ij}} [1 - P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)]^{1-R_{ij}}}. \tag{13}$$

然后, 基于 n_j 个被试在新题 j 上的作答向量 $\boldsymbol{R}_j$ 和每个被试属性掌握模式的后验分布, 假设 n_j 个被试在新题 j 上的作答彼此独立, 可构建对数边际似然函数如下：

$$L(\boldsymbol{q}_j, \boldsymbol{\delta}_j) = \prod_{i=1}^{n_j} \sum_{c=1}^{2^K} Post_{ij}(\boldsymbol{\alpha}_c) \left[\left(R_{ij} \ln P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)\right) + (1 - R_{ij}) \ln\left(1 - P(\boldsymbol{q}_j, \boldsymbol{\delta}_j | \boldsymbol{\alpha}_c)\right) \right]. \tag{14}$$

M 步最大化公式(14)以估计新题 j 的项目参数$\boldsymbol{\delta}_j$。EM 算法依次迭代 E 步和 M 步直到满足预先设定的收敛标准。

3.3 SCADOCM 下 Q 矩阵与项目参数同时标定的基本步骤

SCADOCM 同时标定新题 Q 矩阵和项目参数的具体步骤如下：

步骤 1：新题 q 向量估计。对于新题 j，基于作答了新题 j 的被试在每个属性上的边际掌握概率及其在新题 j 上的作答数据，构建基于 SCAD 的对数似然函数 $SCAD(\boldsymbol{\beta})$，求解 $SCAD(\boldsymbol{\beta})$ 以获得新题 j 的估计 q 向量。

步骤 2：新题项目参数估计。将步骤 1 中的估计 q 向量作为新题 j 的真实 q 向量，基于作答了新题 j 的被试的属性掌握模式后验分布及其在新题 j 上的作答，使用 SCADOCM 中项目参数估计方法估计新题的项目参数。新题 j 标定完成。

步骤 3：对于所有待标定的其他新题，重复步骤 1 和步骤 2 可获得新题的 Q

矩阵估计值和项目参数估计值。直到所有新题标定完成则终止。

4 研究1：模拟题库下SCADOCM的性能验证及与SIE方法的比较研究

研究 1 旨在考查模拟题库下 SCADOCM 在不同标定样本(50、100、500、1000、2000)、属性掌握模式分布(均匀分布、高阶分布、多元正态分布)和项目质量(高质量：$P_j(\mathbf{0})$ (未掌握项目 j 所测量的任一属性的被试在项目 j 上的答对概率)和 1-$P_j(\mathbf{1})$ (掌握项目 j 所测量的所有属性的被试在项目 j 上的答对概率)从 $U(0.05, 0.15)$ 中随机抽取；低质量：$P_j(\mathbf{0})$和 1-$P_j(\mathbf{1})$从$U(0.1, 0.3)$中随机抽取)下标定新题的效果, 并将其与 SIE 方法进行比较。标定样本指作答了新题 j 的被试人数，本文采用陈平和辛涛(2011b)及 Chen 等人(2015)的设定方式即 $n_j = (N×Z)/m$, 其中 N 为参与 CD-CAT 的被试总人数, Z 为每个被试作答新题的个数, m 为待标定的新题个数。本研究共包含 5 (标定样本)×3 (属性掌握模式分布)×2 (项目质量)=30 种模拟实验条件, 每种实验条件重复实验 100 次以减少随机误差。

4.1 数据生成

4.1.1 被试属性掌握模式生成与题库生成

标定样本共 5 个水平, n_j = 50, 100, 500, 1000 和 2000, 被试属性掌握模式分别从均匀分布、高阶分布和多元正态分布 $MVN(0, \sum)$中产生。在均匀分布中, 被试的属性掌握模式从所有可能的属性掌握模式中以均匀的概率产生；在高阶分布中, 被试 i 是否掌握第 k 个属性与被试 i 的一般潜在能力 θ_i 有关, 能力为 θ_i 的被试 i 掌握第 k 个属性的概率为

$$P(\alpha_{ik} = 1|\theta_i, \lambda_{0k}, \lambda_{1k}) = \frac{exp(\lambda_{1k}\theta_i + \lambda_{0k})}{1 + exp(\lambda_{1k}\theta_i + \lambda_{0k})}, \qquad (15)$$

其中，λ_{0k}和 λ_{1k} 为结构参数，研究中设置 $K = 5$，$\boldsymbol{\lambda}_0 = (–1, –0.5, 0, 0.5, 1)$，且对所有属性 k 均有 $\lambda_{1k} = 1.5$，被试 i 的能力值从 $N(0, 1)$中产生(de la Torre & Chiu, 2016)。在 0~1 之间生成一个随机数，将基于上式(公式 15)计算的概率值与随机数进行比较，若概率值大于随机数，被试 i 掌握属性 k，$\alpha_{ik} = 1$，否则被试 i 未掌握属性 k，

$\alpha_{ik}=0$ (Ma & de la Torre, 2020)；在多元正态分布中，属性间的相关设置为 0.5 (J. Chen, 2017; Chiu, 2013)。假设被试 i 的能力向量为$\boldsymbol{\vartheta}_i=(\vartheta_{i1},\cdots,\vartheta_{iK})$，则被试 i 的属性掌握模式$\boldsymbol{\alpha}_i=(\alpha_{i1},\cdots,\alpha_{iK})$可通过以下公式获得(Chiu, 2013)：

$$\alpha_{ik}=\begin{cases}1 & \text{if } \vartheta_{ik}\geqslant \Phi^{-1}\left(\dfrac{k}{K+1}\right)\\ 0 & \text{otherwise}\end{cases}, \tag{16}$$

其中Φ^{-1}是正态分布概率密度的逆函数。

题库生成包含Q矩阵的生成和项目参数的生成。题库中共包含300个题目，每个题目最多测量3个属性，且题库中测量1、2和3个属性的项目均设置为100题。测验测量属性的总个数$K=5$，则共有31种可能的项目q向量，其中测量1个属性的项目q向量个数为5，测量2个属性的项目q向量个数为10，测量3个属性的项目q向量个数也为10。将测量1个属性的5个项目q向量重复20次，测量2个属性的10个项目q向量重复10次，测量3个属性的10个项目q向量重复10次，构成300×5的临时测验Q矩阵。

项目参数的生成如下所示：项目参数$P_j(0)$从$U(0.05, 0.15)$和$U(0.1, 0.3)$中随机抽取，$P_j(1)$从$U(0.85, 0.95)$和$U(0.7, 0.9)$中随机抽取。其他属性掌握模式在新题j上的正确作答概率从$U[P_j(0), P_j(1)]$中随机产生并满足单调性条件，掌握属性个数多的被试在题目j上的答对概率大于掌握属性个数少的被试(de la Torre & Chiu, 2016)。

4.1.2 新题生成

新题生成包括Q矩阵以及项目参数的生成。设置待标定的新题个数$m=20$，新题Q矩阵是大小为20×5的矩阵。从上一步模拟的Q矩阵中随机抽取20行以构建新题Q矩阵，新题项目参数的生成与题库项目参数的生成一致。在生成被试属性掌握模式真值及项目参数真值后，根据给定的认知诊断模型计算被试在每个新题上的正确作答概率，将该正确作答概率与0~1之间的随机数进行比较，如果被试在题目上的正确作答概率大于随机数，则答对题目，否则答错题目。

4.2 CD-CAT 过程及新题标定

研究使用定长终止规则，每个参与测验的被试均作答 20 个旧题和 5 个新题($Z = 5$)。CD-CAT 模拟过程具体如下：

测验开始时对于被试的情况一无所知，因此(1)从题库中随机挑选一个项目作为被试的初始作答题；(2)模拟被试在当前项目上的作答，然后基于被试在已选项目上的作答使用香农熵(shannon entropy, SHE; Cheng, 2009)选题策略为被试从剩余题库中挑选最适合的项目作为其下一个作答项目，重复该步骤直到测验长度达到预先指定的标准。SHE 选题策略理论基础扎实，具有较高的估计精度，已有同时标定新题 Q 矩阵和项目参数的研究也表明 SHE 选题策略下各在线标定方法均具有较好的项目标定精度(Chen et al., 2015; Tan et al., 2022; Zheng & Chang, 2016; 谭青蓉 等, 2021; 张学工, 2010)。因此，研究选用 SHE 作为选题策略；(3)使用极大似然(maximum likelihood estimation, MLE)方法估计被试的属性掌握模式。

在 CD-CAT 模拟过程中，随机从待标定的 20 个新题中抽取 5 个新题并将其置于被试测验过程的随机位置。CD-CAT 测验结束后，基于被试属性边际掌握概率，属性掌握模式后验分布及被试在新题上的作答，分别使用 SCADOCM 和 SIE 方法标定新题的 Q 矩阵和项目参数。

4.3 评价标准

标定效率：即平均运行时间(average running time, ART) ART 用于评估各在线标定方法的标定效率，其计算如下：

$$ART = \frac{\sum_{r=1}^{100} t_r}{100}, \tag{17}$$

其中，t_r 表示第 r 次重复模拟中，各在线标定方法标定新题所用的时间。ART 值越小，说明用于标定新题的方法的效率越高。本文所有实验均在配置为 Intel Core i5-8400 2.81GHz，内存 20G 的计算机上运行，以保证各标定方法的估计效率具有可比性。

属性向量正确估计率(attribute vector correct estimation rate, AVCER) AVCER

用于评估新题 Q 矩阵的估计精度，其计算公式为：

$$AVCER=\frac{1}{100\times m}\sum_{r=1}^{100}\sum_{j=1}^{m}I\left(\hat{\boldsymbol{q}}_j^{(r)}=\boldsymbol{q}_j^{(r)}\right),\tag{18}$$

其中，r 表示 100 次重复模拟实验中的第 r 次重复实验，$\hat{\boldsymbol{q}}_j^{(r)}$ 表示第 r 次重复模拟中新题 j 的 q 向量估计值，$\boldsymbol{q}_j^{(r)}$ 表示第 r 次重复模拟中新题 j 的 q 向量真值。$I\left(\hat{\boldsymbol{q}}_j^{(r)}=\boldsymbol{q}_j^{(r)}\right)$ 为指示性函数，用于评估第 r 次重复模拟中 $\hat{\boldsymbol{q}}_j^{(r)}$ 是否等于 $\boldsymbol{q}_j^{(r)}$。AVCER 值越大，新题 Q 矩阵估计精度越高。

均方根误差(root mean squared error, RMSE) RMSE 指标用于评价新题项目参数的估计精度，其表达式可写为：

$$RMSE=\sqrt{\frac{1}{100\times 2^K\times m}\sum_{r=1}^{100}\sum_{c=1}^{2^K}\sum_{j=1}^{m}\left(\hat{P}_j^{(r)}(\boldsymbol{\alpha}_c)-P_j^{(r)}(\boldsymbol{\alpha}_c)\right)^2},\tag{19}$$

上式中，$\hat{P}_j^{(r)}(\boldsymbol{\alpha}_c)$ 和 $P_j^{(r)}(\boldsymbol{\alpha}_c)$ 分别表示第 r 次重复模拟中属性掌握模式为 $\boldsymbol{\alpha}_c$ 的被试在新题 j 上的正确作答概率估计值和真实值。RMSE 值越小，项目参数的估计精度越高。此外，$P(\mathbf{0})$ 和 1-$P(\mathbf{1})$ 参数的 RMSE 计算公式与公式(19)略有不同，具体如下所示：

$P(\mathbf{0})$:
$$RMSE=\sqrt{\frac{1}{100\times m}\sum_{r=1}^{100}\sum_{j=1}^{m}\left(\hat{P}_j^{(r)}(0)-P_j^{(r)}(0)\right)^2},\tag{20}$$

1-$P(\mathbf{1})$:
$$RMSE=\sqrt{\frac{1}{100\times m}\sum_{r=1}^{100}\sum_{j=1}^{m}\left(\left(1-\hat{P}_j^{(r)}(1)\right)-\left(1-P_j^{(r)}(1)\right)\right)^2}.\tag{21}$$

4.4 研究 1 结果

图 1 至图 3，以及表 1 分别呈现了模拟题库下 SCADOCM 和 SIE 方法的项目标定效率以及项目标定精度结果。各模拟条件下 SCADOCM 的平均运行时间(ART)、属性向量估计正确率(AVCER)以及均方根误差(RMSE)的均值分别为 5.231s、66.4%和 0.101，SIE 方法对应的值分别为 99.893s、0.0%和 0.242。需注意的是，SIE 方法的 AVCER 值均接近于 0.0%，其原因可能在于 SIE 方法中用于估计新题 q 向量的 MLE 方法在 G-DINA 模型下倾向于选择测量所有属性的 q 向量作为新题的估计 q 向量(汪大勋 等, 2020; Chen et al., 2013)。总之，SCADOCM

具有较好的估计效率和项目标定精度，其性能优于 SIE 方法。

图 1 为使用 SCADOCM 和 SIE 方法估计 20 个新题的平均运行时间(单位：秒)。相比于 SCADOCM, SIE 方法的估计效率更低，其所有条件下的平均 ART 值约为 SCADOCM 的 19.095 倍。SCADOCM 和 SIE 的平均 ART 值分别为 5.231s 和 99.893s。在标定样本对各方法标定效率的影响上，SCADOCM 和 SIE 方法的平均运行时间均随标定样本的增加而延长。当标定样为 50 时, SCADOCM 和 SIE 的平均 ART 值分别为 1.216s 和 25.554s，而当标定样本为 2000 时, 2 种方法的平均 ART 值延长至 12.643s 和 222.052s。项目质量对 SCADOCM 和 SIE 的标定效率影响较小。当项目参数范围为 U (0.05, 0.15)和 U (0.1, 0.3)，SCADOCM 的平均 ART 值为 6.543s 和 3.920s，SIE 方法的平均 ART 值为 81.624s 和 118.162s。SCADOCM 的标定效率受属性掌握模式分布的影响较小，SIE 在属性掌握模式为均匀分布和高阶分布下的标定效率略优于正态分布。SCADOCM 和 SIE 的平均 ART 值在属性掌握模式分布为均匀分布时分别为 4.304s 和 58.204s，在属性掌握模式分布为高阶分布时分别为 4.615s 和 65.781s，而在属性掌握模式分布为正态分布时分别为 6.776s 和 175.695s。

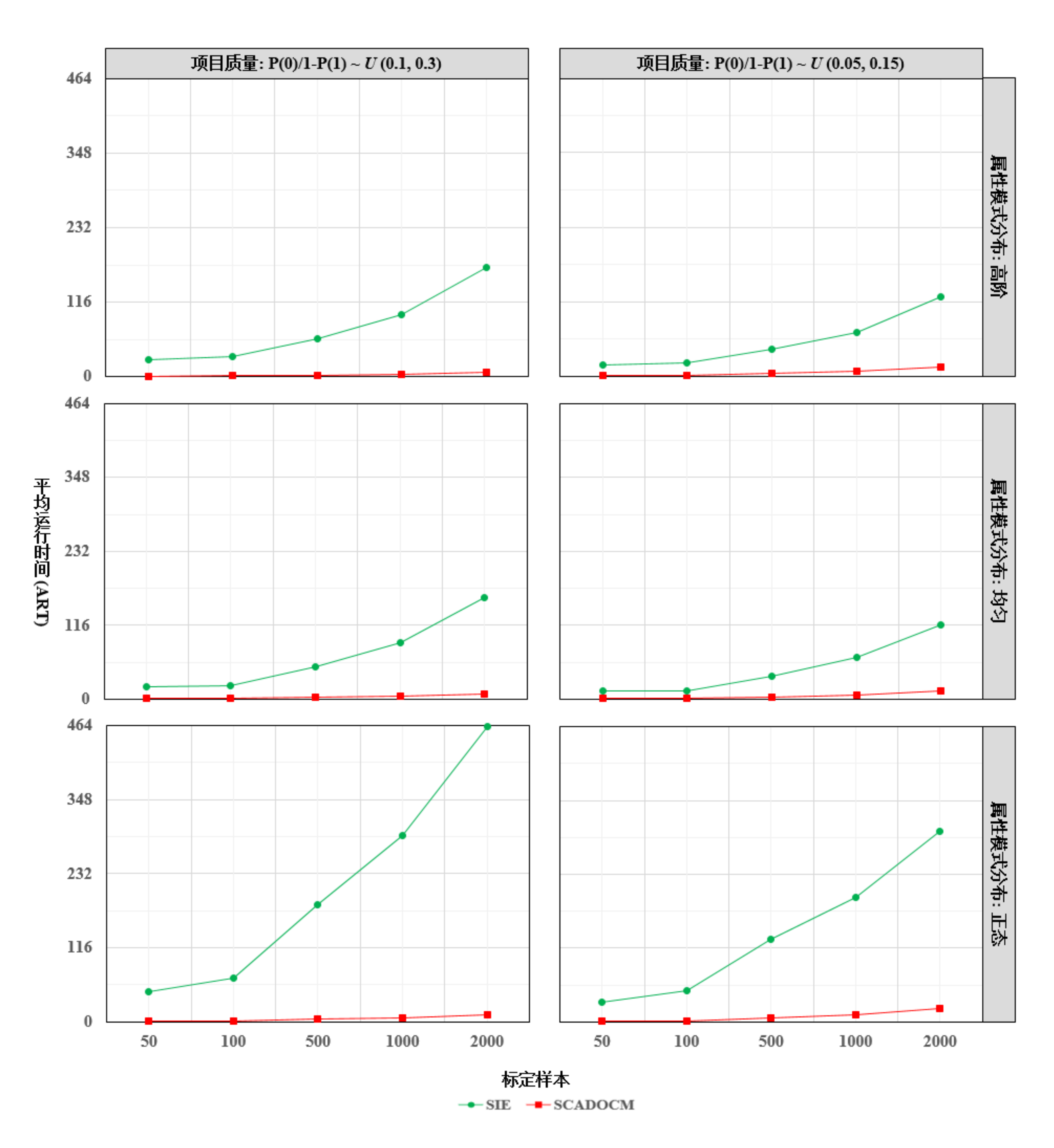


图 1 各在线标定方法在不同条件下的平均运行时间(ART)结果(单位: 秒)

图 2 结果表明, SCADOCM 的 Q 矩阵估计精度高于 SIE 方法, 标定样本、项目质量和属性掌握模式分布均影响 SCADOCM 的 Q 矩阵估计精度，而对 SIE 方法的影响可忽略不计。SIE 方法在各模拟条件下的 AVCER 值均接近于 0。SCADOCM 的 Q 矩阵估计精度随标定样本的增加而提高。各标定样本(50、100、500、1000 和 2000)下，SCADOCM 的 AVCER 均值分别为：38.3%、48.9%、74.5%、82.3%和 88.3%。在标定样本达到一定的数量后，样本量对 SCADOCM 的 Q 矩阵估计精度的影响逐渐减小。当标定样本从 50 增加到 100 时，SCADOCM 的 AVCER 指标差值为 10.6%，从 100 增加到 500 时，SCADOCM 的 AVCER 差值为 25.6%，每增加 50 个被试所增加的 AVCER 值平均为 3.2%，而从 1000 增加到 2000 时，SCADOCM 的 AVCER 差值仅为 6.0%，每增加 50 个被试所增加的 AVCER 值平均为 0.3%。项目质量越高，SCADOCM 的 Q 矩阵估计精度越高，当项目参数范围从 U (0.05, 0.15)变化到 U (0.1, 0.3)时，AVCER 值在固定标定样本和属性掌握模式分布下单调递减。在项目参数范围为 U (0.05, 0.15)时，SCADOCM 的 AVCER 值在 40.4%~96.0%之间，项目参数范围为 U (0.1, 0.3)时，SCADOCM 的 AVCER 值在 30.2%~89.4%之间。在属性掌握模式分布对 Q 矩阵标定精度的影响上, 多数实验条件下，SCADOCM 的 Q 矩阵估计精度在属性掌握模式为均匀分布时最好, 高阶分布时次之, 正态分布时最差。其可能的原因在于，均匀分布下每种属性掌握模式的被试人数都较为均匀，而高阶分布和正态分布下某些属性掌握模式的被试人数非常少，尤其是正态分布下某些属性掌握模式的被试人数更少，这不利于正确 q 向量的识别(Chiu, 2013; Wang et al., 2018)，从而导致高阶分布和正态分布下的 Q 矩阵估计精度更低。SCADOCM 在均匀、高阶和正态分布下的 Q 矩阵估计精度范围分别为 35.2%~96.0%, 33.7%~93.4%和 30.2%~86.0%。但在项目参数范围为 U (0.05, 0.15)，标定样本为 100 时，SCADOCM 的 AVCER 值在高阶分布下更大。此条件下，SCADOCM 在均匀和高阶分布下的 AVCER 值分别为 58.4%和 59.9%。

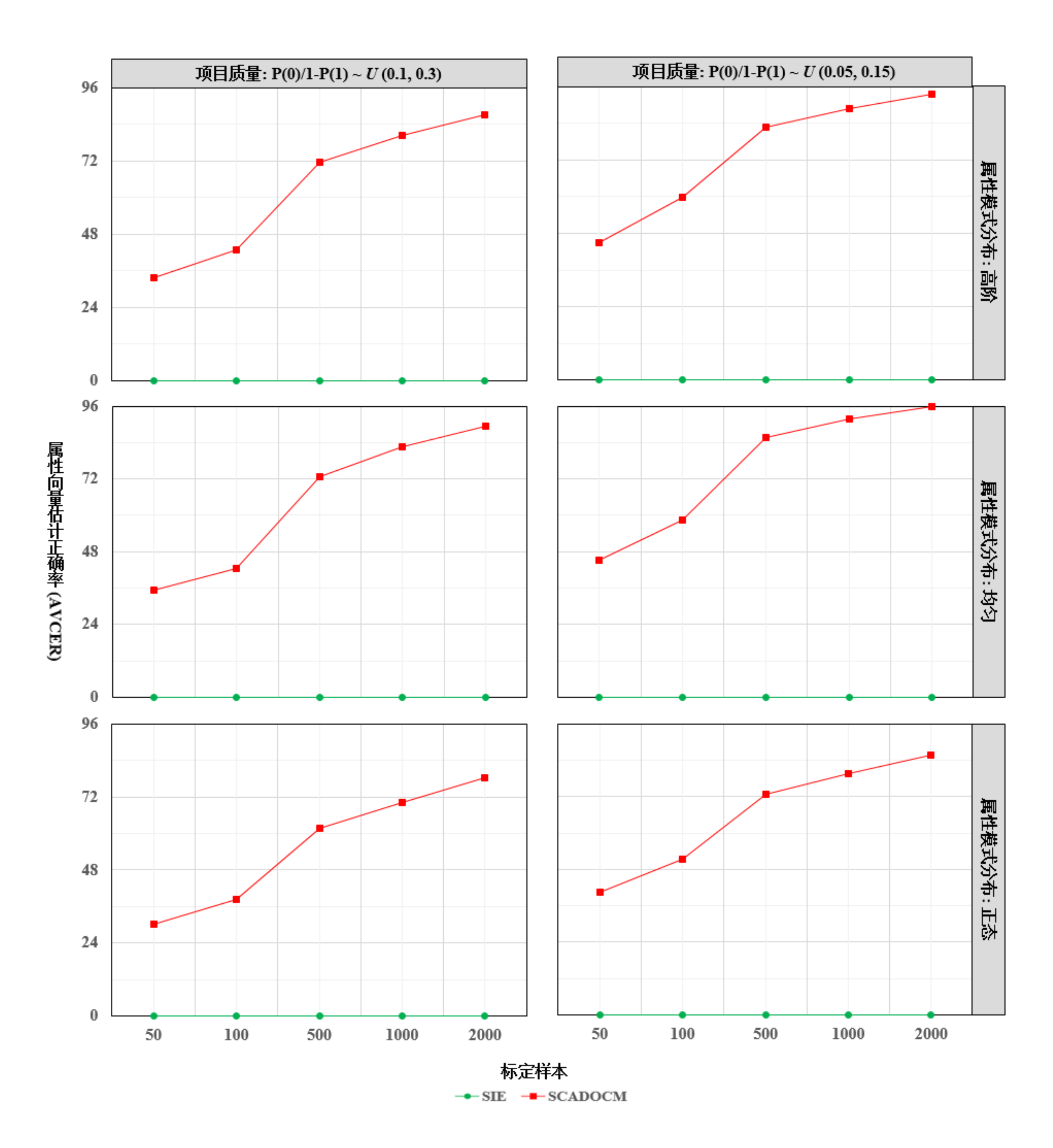


图 2 各在线标定方法在不同条件下的属性向量估计正确率(AVCER)结果

图 3 为 SCADOCM 和 SIE 的项目参数标定结果。SCADOCM 的项目参数标定精度优于 SIE 方法，两方法均受标定样本、项目质量和属性掌握模式分布的影响。随着标定样本的增加，SCADOCM 和 SIE 方法的项目参数标定精度提高。各标定样本下，SCADOCM 的平均 RMSE 值分别为 0.188、0.145、0.076、0.057 和 0.042，SIE 的平均 RMSE 值分别为 0.400、0.337、0.200、0.156 和 0.120。SCADOCM 和 SIE 在标定样本为 50 和 2000 之间的平均 RMSE 指标差值分别为 0.146 和 0.280。标定样本对 SIE 方法的影响略大于 SCADOCM。SCADOCM 和 SIE 的项目参数标定精度在部分实验条件下随项目质量的提升而略有升高，但在部分实验条件下随项目质量的提升而略有下降。总体上来说，SCADOCM 在两项目参数范围下(U (0.05, 0.15)和 U (0.1, 0.3))的平均 RMSE 值分别为 0.101 (0.020~0.231)和 0.102 (0.025~0.220)，平均 RMSE 值变大，SIE 在两项目参数下的平均 RMSE 值分别为 0.235 (0.046~0.448)和 0.250 (0.058~0.429)，平均 RMSE 值变大。在属性掌握模式分布为正态分布时，SCADOCM 在项目参数范围为 U (0.05, 0.15)时具有更大的 RMSE 值，两项目参数范围间的 RMSE 最大差值为 0.013；在属性掌握模式为正态分布且标定样本为 50 和 100 时，SIE 在项目参数范围为 U (0.05, 0.15)时具有更大的 RMSE 值，两项目参数范围间的 RMSE 差值在标定样本为 50 时为 0.019。这可能是标定样本和属性掌握模式分布相互作用的结果。新题的项目参数标定精度在标定样本量少的情况下较低，而在标定样本少且属性掌握模式分布为正态分布时，更有可能出现某些属性掌握模式下的被试数量多而另一些属性掌握模式下的被试缺失的情况，两者共同作用可能导致项目质量高时的 RMSE 值略大于项目质量低时，但是这种差异是较小的，且可以通过增大样本量或改变属性掌握模式分布扭转这种趋势。在属性掌握模式分布对项目参数标定精度的影响上, SCADOCM 和 SIE 方法的项目参数标定精度在属性掌握模式为均匀分布时最好, 高阶分布时次之, 正态分布时最差。均匀、高阶和正态分布下，SCADOCM 的 RMSE 范围分别为 0.020~0.154、0.028~0.185 和 0.070~0.231，SIE 的 RMSE 范围分别为 0.046~0.378、0.079~0.403 和 0.221~0.448。

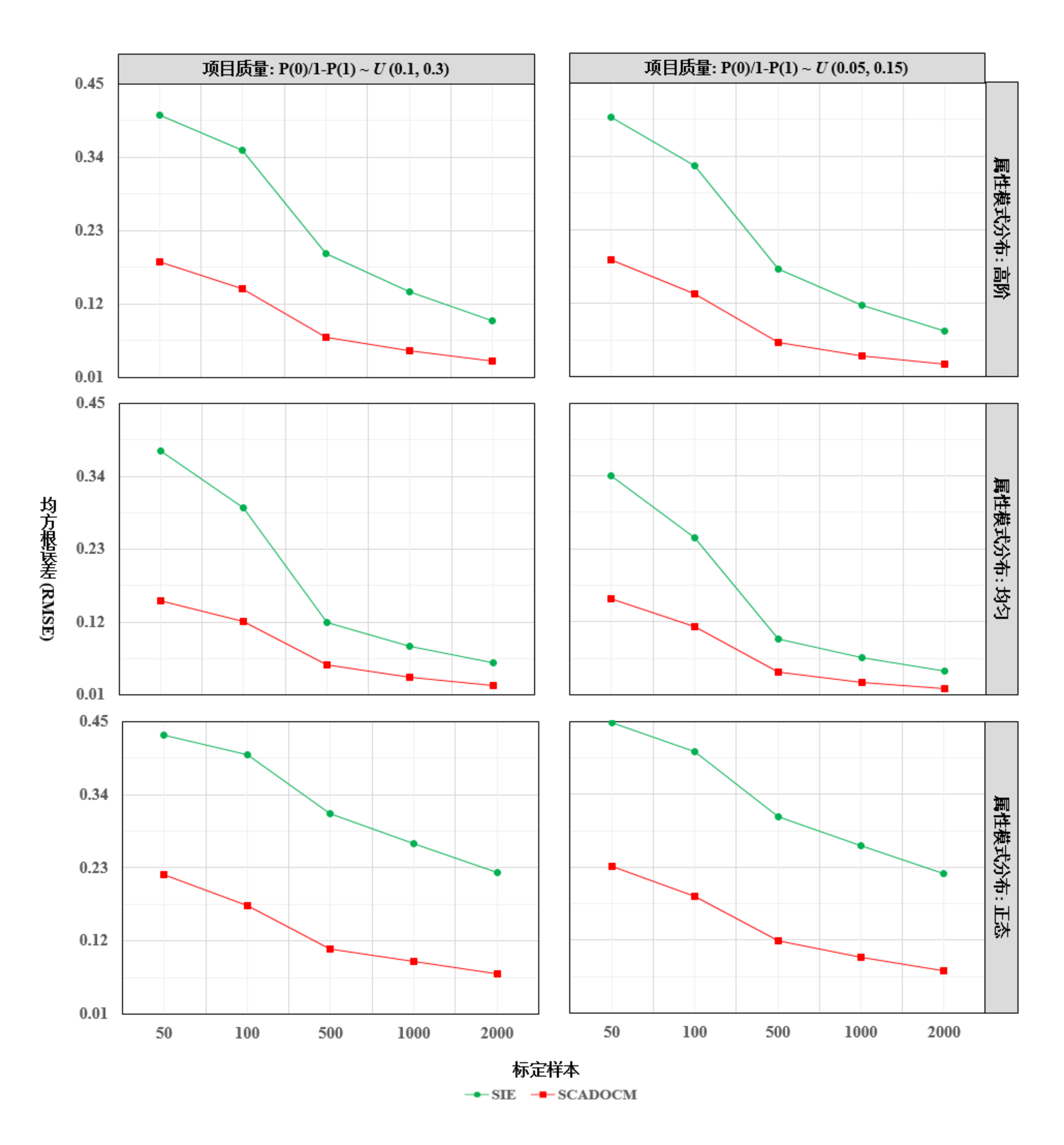


图 3 各在线标定方法在不同条件下项目参数标定精度(RMSE)结果

表 1 呈现了 SIE 和 SCADOCM 的 *P*(**0**)和 1-*P*(**1**)参数标定结果。结果表明 SCADOCM 在 *P*(**0**)和 1-*P*(**1**)参数上具有较好的标定精度，优于 SIE 方法，尤其在

标定样本量少的情况下。SIE 和 SCADOCM 均受标定样本、项目质量和属性掌握模式分布的影响。SIE 和 SCADOCM 的 $P(\mathbf{0})$和 1-$P(\mathbf{1})$参数标定精度随着标定样本的增加而提高。对于 $P(\mathbf{0})$参数，SIE 在各标定样本下的平均 RMSE 值分别为 0.223、0.155、0.066、0.046 和 0.032，SCADOCM 对应的平均 RMSE 值分别为 0.155、0.120、0.048、0.032 和 0.022；对于 1-$P(\mathbf{1})$参数，SIE 在各标定样本下的平均 RMSE 值分别为 0.235、0.163、0.067、0.046 和 0.033，SCADOCM 对应的平均 RMSE 值分别为 0.118、0.087、0.037、0.026 和 0.018。SIE 和 SCADOCM 在 $P(\mathbf{0})$和 1-$P(\mathbf{1})$参数上的标定精度随项目质量的提升而升高，除标定样本为 50 的情况。标定样本为 50 时，SCADOCM 在项目质量低时的标定精度高于项目质量高时，但 RMSE 差值较小，最大差值为 0.022。在属性掌握模式分布对 $P(\mathbf{0})$和 1-$P(\mathbf{1})$参数标定精度的影响上, SIE 和 SCADOCM 在属性掌握模式为高阶分布时的 $P(\mathbf{0})$和 1-$P(\mathbf{1})$参数标定精度略优于均匀分布和正态分布。对于 $P(\mathbf{0})$参数，均匀、高阶和正态分布下，SIE 的 RMSE 范围分别为 0.038~0.362、0.019~0.180 和 0.023~0.229，SCADOCM 的 RMSE 范围分别为 0.018~0.184、0.014~0.133 和 0.019~0.161；对于 1-$P(\mathbf{1})$参数, 均匀、高阶和正态分布下，SIE 的 RMSE 范围分别为 0.039~0.356、0.019~0.186 和 0.023~0.232，SCADOCM 的 RMSE 范围分别为 0.015~0.122、0.013~0.107 和 0.017~0.134。

表 1 各在线标定方法在不同条件下 $P(\mathbf{0})$和 1- $P(\mathbf{1})$参数标定精度(RMSE)结果

项目质量	属性模式分布	标定样本	$P(\mathbf{0})$		1- $P(\mathbf{1})$	
			SIE	SCADOCM	SIE	SCADOCM
0.1-0.3	高阶	50	0.180	0.133	0.186	0.107
		100	0.122	0.108	0.127	0.085
		500	0.055	0.045	0.057	0.037
		1000	0.037	0.031	0.039	0.026
		2000	0.027	0.022	0.028	0.018
	均匀	50	0.362	0.162	0.356	0.122
		100	0.260	0.148	0.281	0.099
		500	0.111	0.068	0.113	0.046
		1000	0.079	0.041	0.078	0.032
		2000	0.053	0.027	0.054	0.022
	正态	50	0.229	0.160	0.232	0.134
		100	0.154	0.124	0.155	0.101
		500	0.065	0.058	0.065	0.045
		1000	0.046	0.041	0.047	0.033

		2000	0.033	0.030	0.034	0.024
	高阶	50	0.131	0.127	0.127	0.095
		100	0.086	0.088	0.088	0.066
		500	0.038	0.033	0.038	0.027
		1000	0.026	0.021	0.026	0.019
		2000	0.019	0.014	0.019	0.013
0.05-0.15	均匀	50	0.269	0.184	0.329	0.122
		100	0.198	0.142	0.218	0.087
		500	0.079	0.041	0.079	0.034
		1000	0.057	0.027	0.055	0.023
		2000	0.038	0.018	0.039	0.015
	正态	50	0.169	0.161	0.177	0.125
		100	0.107	0.107	0.110	0.084
		500	0.046	0.044	0.047	0.035
		1000	0.033	0.029	0.033	0.025
		2000	0.023	0.019	0.023	0.017

5 研究2：真实题库下SCADOCM的性能验证

基于研究一的结果，考虑到 SIE 方法在各实验条件下的 Q 矩阵标定精度均较低，不适用 G-DINA 等模型。因此，研究 2 仅考查真实题库下 SCADOCM 在不同标定样本(50、100、500、1000、2000)和属性掌握模式分布(均匀分布、高阶分布、多元正态分布)下标定新题的效果。本研究共包含 5 (标定样本)×3 (属性掌握模式分布)=15 种模拟实验条件，每种实验条件重复实验 100 次以减少随机误差。

5.1 真实题库及新题指定

真实题库：因可为患者提供全面且详细的症状图谱等独有的优势，认知诊断在心理障碍评估与诊断中的应用日益增加。如研究者将认知诊断应用于病理性赌博、分裂型人格、边缘型人格、焦虑、抑郁和网络成瘾等的评估与诊断(de la Torre et al., 2018; Peng et al., 2019; Templin & Henson, 2006; Tu et al., 2017; Xi et al., 2020; 史双双, 2017)。史双双(2017)基于《精神障碍诊断与统计手册》第五版(DSM-5)中定义的网络成瘾症状标准构建了网络成瘾题库，且在实践中已验证该网络成瘾题库的信效度等均符合心理测量学要求。本实验使用该网络成瘾题库作为真实题库，题库中包含 263 道二级计分项目，每个项目最多测量 3 个属性(症状标准)，共测量了 9 个属性(如表 2 所示)。根据 DSM-5 的诊断标准，被试满足

9 个症状标准中的 5 个或 5 个以上便可诊断为网络成瘾。实验使用史双双(2017)研究中的原始 Q 矩阵作为真实 Q 矩阵，并基于该真实 Q 矩阵以及 1558 个真实被试的作答数据使用 G-DINA 模型估计题库的项目参数，项目参数描述性统计结果如表 3 所示。另外，题库中所有项目的参数结果如附表 1 所示。选用 G-DINA 模型进行分析，主要考虑到 G-DINA 模型既允许属性间存在补偿关系，也允许属性间存在非补偿关系，适合于网络成瘾测验的分析，且模型-资料拟合检验(表 4)结果发现 G-DINA 模型较 DINA 等其它约束的认知诊断模型更能拟合该网络成瘾数据。

新题指定：从网络成瘾题库中随机抽取 20 个项目作为待标定 Q 矩阵与项目参数的新题。

研究 2 中被试属性掌握模式的生成，CD-CAT 过程与新题标定以及评价标准均与研究 1 保持一致。需注意的是，研究 2 中项目参数“真值”是基于已有研究中给定的由专家标定的 Q 矩阵和所有被试的真实作答数据使用 G-DINA 模型估计的结果，基于该“真值”计算的 RMSE 指标反映的是项目参数估计结果之间的一致性。

表 2 DSM-5 中定义的网络成瘾症状标准

ID	症状标准
A1	沉迷于网络游戏(如，重温过去的游戏经历或期望下一次游戏，网络游戏成为日常的主导活动)。
A2	远离网络游戏时出现戒断症状(如，易怒、焦虑或悲伤，但没有药物戒断的身体迹象)。
A3	耐受性——需要花更多的时间参与网络游戏。
A4	试图控制网络游戏的参与不成功。
A5	因网络游戏而对以前的爱好和娱乐失去兴趣，但网络游戏除外。
A6	尽管了解心理社会问题，但仍继续过度使用网络游戏。
A7	向家庭成员、治疗师或者其他人撒谎参与网络游戏的次数。
A8	利用网络游戏来逃避或缓解消极情绪(如，无助感、焦虑、内疚)
A9	因参与网络游戏而危及或失去重要的人际关系、工作、教育或职业机会。

表 3 网络成瘾题库项目参数的描述性统计

项目参数	最小值	最大值	平均值	标准差
1-*P*(**1**)	0.161	0.500	0.450	0.072
P(**0**)	0.004	0.500	0.069	0.082

注：*P*(**0**)指未掌握项目所测量的任一属性的被试在项目上的答对概率，*P*(**1**)指掌握项目所测量的所有属性的被试在项目上的答对概率。

表 4 网络成瘾题库模型-资料拟合检验结果

模型	AIC	BIC	LL
DINA	309348.5428	314897.6939	-153637.2714
DINO	309803.4409	315352.5920	-153864.7204
ACDM	307764.2211	313586.2812	-152794.1105
G-DINA	307426.2025	313574.6833	-152564.1012

5.2 研究 2 结果

表 3 呈现了网络成瘾题库项目参数的描述性统计，相比研究 1 模拟题库中项目的质量(*P*(**0**)/(1-*P*(**1**))~$U(0.05, 0.15)$和$U(0.1, 0.3)$)，网络成瘾题库中项目的质量更低。在该真实题库下进一步验证 SCADOCM 的性能，可以进一步考察 SCADOCM 的适用范围以及该方法在实践中应用时的稳健性。

表 5 呈现了网络成瘾题库下新方法 SCADOCM 的项目标定效率、*Q* 矩阵估计精度和项目参数标定一致性结果。结果表明，真实题库下 SCADOCM 仍具有较好的估计效率、*Q* 矩阵估计精度和项目参数标定一致性。具体而言，各模拟条件下 SCADOCM 的 ART、AVCER 以及 RMSE 的均值分别为 37.612s、79.8%和 0.101。

使用 SCADOCM 估计 20 个新题的平均运行时间(单位：秒)如表 5 所示。SCADOCM 的平均 ART 值为 37.612s。在标定样本对 SCADOCM 标定效率的影响上，SCADOCM 的平均运行时间均随标定样本的增加而延长。当标定样为 50 时, SCADOCM 的平均 ART 值为 4.507s；而当标定样本为 2000 时, 其平均 ART 值延长至 101.849s。SCADOCM 的标定效率在各属性掌握模式分布之间的差异不

大。SCADOCM 的平均 ART 值在属性掌握模式分布为均匀分布、高阶分布和正态分布时分别为 37.567s、38.060s 和 37.209s。

表 5 结果表明，标定样本和属性掌握模式分布均影响 SCADOCM 的 Q 矩阵估计精度。SCADOCM 的 Q 矩阵估计精度随标定样本的增加而提高。各标定样本(50、100、500、1000 和 2000)下，SCADOCM 的 AVCER 均值分别为：57.0%、69.8%、88.0%、91.2%和 92.8%。与模拟题库一致，在标定样本达到一定的数量后，样本量对 SCADOCM 的 Q 矩阵估计精度的影响逐渐减小。当标定样本从 50 增加到 100 时，SCADOCM 的 AVCER 指标差值为 12.8%，从 100 增加到 500 时，SCADOCM 的 AVCER 差值为 18.2%，每增加 50 个被试所增加的 AVCER 值平均为 2.3%，而从 1000 增加到 2000 时，SCADOCM 的 AVCER 差值仅为 1.6%，每增加 50 个被试所增加的 AVCER 值平均为 0.1%。在属性掌握模式分布对 Q 矩阵标定精度的影响上, SCADOCM 的 Q 矩阵估计精度在属性掌握模式为均匀分布时最好, 高阶分布时次之, 正态分布时最差。SCADOCM 在均匀、高阶和正态分布下的 Q 矩阵估计精度范围分别为 69.7%~97.8%, 56.0%~94.5%和 45.4%~86.3%。

与模拟题库一致，SCADOCM 的项目参数标定一致性受标定样本和属性掌握模式分布的影响。随着标定样本的增加，SCADOCM 的项目参数标定一致性提高。各标定样本下，SCADOCM 的平均 RMSE 值分别为 0.192、0.135、0.069、0.058 和 0.052。在属性掌握模式分布对项目参数标定一致性的影响上，SCADOCM 的项目参数标定一致性在属性掌握模式为均匀分布时最好, 高阶分布时次之, 正态分布时最差。均匀、高阶和正态分布下，SCADOCM 的 RMSE 范围分别为 0.019~0.142、0.032~0.189 和 0.105~0.244。

表 5 真实题库下 SCADOCM 的新题标定结果

评价指标	标定样本 / 分布	50	100	500	1000	2000
ART (单位：秒)	均匀	4.585	6.954	25.610	49.539	101.146
	高阶	4.325	6.739	26.217	49.898	103.118
	正态	4.612	6.946	25.035	48.168	101.284
AVCER	均匀	0.697	0.782	0.943	0.968	0.978
	高阶	0.560	0.702	0.882	0.924	0.945
	正态	0.454	0.611	0.815	0.845	0.863
RMSE	均匀	0.142	0.093	0.037	0.026	0.019
	高阶	0.189	0.125	0.053	0.040	0.032
	正态	0.244	0.187	0.118	0.109	0.105

6 讨论与未来研究方向

如何才能使已构建好的 CD-CAT 在实际测验中长久有效地发挥作用，高效地为测验使用者提供准确详尽的诊断结果？行之有效的题库维护或更新方法是必不可少的。项目增补对于题库维护起着至关重要的作用，而在线标定是一种有效的项目增补方法。然而，CD-CAT 中有关 Q 矩阵与项目参数同时性在线标定方法的研究较少，且基本是基于 DINA 模型提出。而 G-DINA 模型下有关 Q 矩阵与项目参数同时性在线标定方法的研究几乎空白，这一定程度上有碍于 CD-CAT 在实际测验中的进一步推广。

本研究基于正则化方法选择特征的思路，提出了适用于 G-DINA 等模型的在线标定新方法 SCADOCM，以期为 CD-CAT 题库的项目增补提供新的方法支持。新方法 SCADOCM 使用正则化方法标定新题的 Q 矩阵，相比已有在线标定方法中所使用的最优子集思路，可有效节约新题标定的时间，为 CD-CAT 中 Q 矩阵与项目参数同时性在线标定方法的研究提供了新的思路与视角。通过模拟与真实题库下的 Monte Carlo 模拟研究检验 SCADOCM 的可行性与合理性，考察标定样本、项目质量以及属性掌握模式分布等因素对其性能的影响，并与传统的 SIE 方法进行比较。研究结果表明，新方法 SCADOCM 在各模拟条件下都具有较为理想的标定效率和标定精度，且优于 SIE 方法。如，模拟题库下 SIE 的平均 ART 值是 SCADOCM 的 19.096 倍，说明 SCADOCM 具有更高的标定效率。SCADOCM 的平均 AVCER 值比 SIE 高 66.4%，且 SCADOCM 的平均 RMSE 值比 SIE 低 0.141，显示 SCADOCM 在标定精度上表现出更好的性能。另外，研究结果显示，SIE 的 Q 矩阵估计精度在各条件下几乎都接近于 0。其可能的原因在于：研究中所用评估 Q 矩阵估计精度的 AVCER 指标，评估题目的整个估计 q 向量和真实 q 向量之间的一致性，也即 q 向量模式的估计精度。SIE 方法中使用 MLE 方法估计新题 q 向量，而在 G-DINA 模型下，MLE 方法倾向于选择测量所有属性的 q 向量(即全为 1 的 q 向量)作为新题的估计 q 向量(汪大勋 等, 2020; Chen et al., 2013)。例如，测验测量属性个数 $K = 5$ 时，SIE 方法选择 q 向量 $\boldsymbol{q}$=[1 1 1 1 1]作为题目的估计 q 向量，实验结果调查也证实了这一点。在模拟实验中，设置测验共测量 5 个属性，每个题目(旧题和新题)最多测量 3 个属性，使用 SIE 标定新题 Q 矩阵偏

向于指定每个题目都测量 5 个属性，此时新题 Q 矩阵的属性向量估计精度低于随机分配概率，出现 AVCER 在 0 左右的结果。假设 20 个新题均测量 3 个属性，则 20×5 的新题 Q 矩阵中有 60 个元素为 1，40 个元素为 0，此时 SIE 方法的属性估计精度约为 60%，也即 SIE 方法的属性估计精度最大值为 60%；研究中 20 个新题的 q 向量从 300 个旧题(测量 1、2 和 3 个属性的项目均为 100 题)中随机抽取，20×5 的新题 Q 矩阵中元素为 1 的个数大多数情况下小于 50 个，该类情况下 SIE 方法的属性估计精度低于 50%。研究 1 中各模拟条件下 SIE 方法的平均属性估计精度为 39.8%，大于 0，低于 50%。研究 1 在 SIE 方法的 AVCER 极低的情况下仍保留了该方法作为比较基准，主要考虑到该结果可以为其他研究者和实践者提供参考与借鉴，他们未来在 G-DINA 等饱和模型下进行在线标定方法研究时可以避免选择该方法作为比较基准。此外，SIE 方法标定新题 Q 矩阵时未考虑模型复杂性，可能不适用于 G-DINA 等饱和模型，可以从对模型复杂性进行惩罚这一思路入手改进该方法。具体来说，使用 SIE 标定新题 Q 矩阵时，基于模型复杂性的考虑，对似然进行惩罚，构建 BIC 指标，选择能使 BIC 值最小的 q 向量作为新题的估计 q 向量。初步的预实验表明：改进的 SIE 方法的项目标定精度优于 SIE 方法。项目参数 $P(\mathbf{0})$和 1-$P(\mathbf{1})$的取值范围为 $U(0.1, 0.3)$，属性掌握模式分布为正态分布，标定样本为 500 时，改进 SIE 方法的平均运行时间(ART)、属性向量正确估计率(AVCER)、项目参数均方根误差(RMSE)、$P(\mathbf{0})$和 1-$P(\mathbf{1})$参数的 RMSE 值分别为 153.758s、54.9%、0.104、0.058 和 0.048，Q 矩阵标定精度远优于 SIE 方法，但仍不如新方法 SCADOCM(此条件下 SCADOCM 的 AVCER 值为 61.7%)。

尽管研究是针对 CD-CAT 题库开发与维护过程中项目增补的技术难点，开发高效可行的在线标定方法，但其与心理学问题是紧密相关的。心理测量学是研究心理学的工具，心理问题(如抑郁、焦虑)的评估与测量都离不开心理测量学。CD-CAT 作为一种新的测验形式，可以更高效、精准地筛查存在心理问题的患者，缓解患者(如抑郁症、躁狂症)做包含大量题目的问卷时的痛苦，减轻其测试的负担。更为重要的是，CD-CAT 可以帮助测验使用者了解患者在某种心理问题各个症状上的表现，更快地获得诊断结果，且能依据该诊断结果制定针对性的治疗方案。在心理测评中应用 CD-CAT 对患者和测验使用者都具有重要的意义，研究

致力于解决 CD-CAT 在实际测验中持续应用时所面临的一大挑战，也即 CD-CAT 题库构建与维护过程中进行项目增补所需应对的技术难题，促进 CD-CAT 在心理测评实践中的应用与推广，以期帮助测验使用者获得更为精细的诊断结果，制定相应的治疗计划，这与心理学问题息息相关。

虽然研究丰富了 CD-CAT 中有关在线标定方法的研究，但仍有许多有待进一步完善及深入研究的地方。具体分述如下：

第一，新方法 SCADOCM 中使用 SCAD 来标定新题的 *Q* 矩阵，其性能受λ参数影响，一个合适且优良的λ值可提高 SCADOCM 的 *Q* 矩阵标定精度，进而提高该方法的项目标定精度(Fan & Li, 2001; Fan & Lv, 2010; Fan & Tang, 2013; Zhang et al., 2010)。研究使用数据挖掘领域中比较常用且效果较好的 BIC 准则来选择λ值(Wang et al., 2007; Zhang et al., 2010)，尽管研究表明 SCADOCM 中使用该准则选择λ值时可获得令人满意的项目标定精度，但在 *Q* 矩阵与项目参数同时性在线标定方法研究中是否存在更好的λ参数选择准则仍是一个值得探讨的问题。未来研究中可对已有的λ参数选择准则进行系统比较，以为 SCADOCM 中λ参数的选择提供建议与参考。

第二，本研究仅考虑了定长的 CD-CAT 终止规则，但变长终止规则更好地体现了 CD-CAT 的自适应特征。变长终止规则下如何实现新题的标定是未来研究中可以进一步讨论和探索的。例如，在变长终止规则下应如何为考生分配新题，新题的分配方式是否会影响最终的项目标定精度等。此外，本研究的研究设计围绕在线标定方法的性能检验及相关因素对其的影响展开，尚未探索测量不变性的问题。不同于以往研究中被试作答矩阵完整，题目 *Q* 矩阵已知且正确的情况(Bradshaw & Madsion, 2015; de la Torre & Lee, 2010; Madsion & Bradshaw, 2018)，CD-CAT 中同时标定新题 *Q* 矩阵和项目参数时，被试的作答矩阵是一个缺乏较多作答数据的稀疏矩阵，每个题目都只有部分被试作答，每个被试也只作答少数几个题目(若被试需作答的待标定新题过多，CD-CAT 的测验长度可能大幅增加，加重被试的作答负担)，且题目 *Q* 矩阵未知。此时，即使标定样本大(如 1000 人)，项目参数的标定精度也较低，无法保证测量不变性。Bradshaw 和 Madsion (2015)

在其研究中指出，在参数估计精度较低的情况下，很难观察到较强的测量不变性，其在研究中也提到，模型数据拟合假设以其它形式违背(如 Q 矩阵错误指定，Bradshaw & Madsion, 2015)时，可能也会影响被试的分类一致性。因此，在被试作答矩阵为稀疏矩阵，Q 矩阵未知或指定错误的情况下，是否仍能观察到测量不变性，在何种条件下可以观察到测量不变性是未来研究可以考虑的一个方向。

第三，CD-CAT 中已有的 Q 矩阵与项目参数同时性在线标定方法重点关注被试的作答数据，而忽视了在计算机化测验中可以便捷获取的过程性数据，如作答反应时间(response times, RTs)数据。以往研究表明，反应时间数据可以提供有关被试认知过程的极具价值的信息，其能提高项目参数的估计精度(Kang et al., 2020; Klein Entink et al., 2009; van der Linden et al., 2010)。未来研究可考虑在作答数据与反应时间数据的联合框架内标定新题，以检验反应时间数据是否有助于提高在线标定方法的标定精度。

第四，研究假设 CD-CAT 题库测量的属性个数是固定且已知的，但在 CD-CAT 的持续使用过程中可能会不定时的往题库中增加新的属性。毫无疑问，各在线标定方法的性能会随新属性的增加而有所波动，在测验测量属性个数随时间发生变化的情况下如何提高 CD-CAT 中已有 Q 矩阵与项目参数同时性在线标定方法的性能是研究者所面临的一大挑战。另外，研究假设测验属性间相互独立，在属性间存在层级关系(如，线型、分支型、收敛型等)时，各在线标定方法的性能如何仍有待于探索。

第五，本文不仅在模拟题库下检验了各在线标定方法的性能，还进一步在真实题库下验证了 SCADOCM 方法的性能，保证了研究的生态性。研究结果表明 SCADOCM 方法的标定性能在模拟题库和真实题库下均较为理想，SCADOCM 方法的可推广性较好，可以为实践应用提供一定的指导。但与以往国内外项目参数同时性在线标定方法的研究(Chen et al., 2015; Tan et al., 2022; 陈平, 辛涛, 2011b; 谭青蓉 等, 2021)一致，研究使用的始终是 Monte Carlo 模拟方法，并未在实证研究情境中加以应用，评估其性能。主要原因在于：在真实测验情境中验证在线标定方法的性能，需要事先构建好一个可以用于实际测验的真实 CD-CAT 测试平台，这需要耗费大量的时间和精力，目前这种平台较难获取。这是本研究，甚至于目前 CD-CAT 中在线标定研究的不足之处，也是未来可进一步深入的研

究方向。总之, CD-CAT 中 Q 矩阵与项目参数同时性在线标定方法的研究仍有待进一步深化。

7 结论

研究主要结论如下：

(1) SCADOCM 具备较好的项目标定性能，优于 SIE 方法。此外，SIE 的 Q 矩阵估计精度在各条件下几乎都接近于 0，该方法不适用于 G-DINA 等饱和模型。

(2) 整体而言，SCADOCM 和 SIE 在标定样本大、项目质量高、属性掌握模式分布为均匀分布和高阶分布时的项目标定精度比标定样本小、项目质量低、属性掌握模式分布为正态分布时更高。

(3) SCADOCM 在标定样本少时的项目标定效率更高，项目质量和属性掌握模式分布对其标定效率的影响较小。SIE 方法在标定样本少时的标定效率比标定样本大时更高，在属性掌握模式分布为均匀分布和高阶分布时的标定效率比属性掌握模式分布为正态分布时更高，其标定效率受项目质量的影响较小。

参考文献

Ban, J. C., Hanson, B. A., Wang, T., Yi, Q., & Harris, D. J. (2001). A comparative study of on-line pretest item—calibration/scaling methods in computerized adaptive testing. *Journal of Educational Measurement*, *38*(3), 191–212.

Bradshaw, L. P., & Madison, M. J. (2015). Invariance properties for general diagnostic classification models. *International Journal of Testing*, *16*(2), 99–118.

Breheny, P., & Huang, J. (2011). Coordinate descent algorithms for nonconvex penalized regression, with applications to biological feature selection. *The annals of applied statistics*, *5*(1), 232–353.

Chen, J. (2017). A residual–based approach to validate Q–matrix specifications. *Applied Psychological Measurement*, *41*(4), 277–293.

Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, *50*(2), 123–

140.

Chen, P. (2016). Two new online calibration methods for computerized adaptive testing. *Acta Psychologica Sinica*, *48*(9), 1184–1198.

[陈平. (2016). 两种新的计算机化自适应测验在线标定方法. *心理学报*, *48*(9), 1184–1198.]

Chen, P. (2017). A comparative study of online item calibration methods in multidimensional computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *42*(5), 559–590.

Chen, P., & Wang, C. (2015). A new online calibration method for multidimensional computerized adaptive testing. *Psychometrika*, *81*(3), 674–701.

Chen, P., Wang, C., Xin, T., & Chang, H. H. (2017). Developing new online calibration methods for multidimensional computerized adaptive testing. *British Journal of Mathematical & Statistical Psychology*, *70*(1), 81–117.

Chen, P., & Xin, T. (2011a). Developing on–line calibration methods for cognitive diagnostic computerized adaptive testing. *Acta Psychologica Sinica*, *43*(6), 710–724.

[陈平, 辛涛. (2011a). 认知诊断计算机化自适应测验中在线标定方法的开发. *心理学报*, *43*(6), 710–724.]

Chen, P., & Xin, T. (2011b). Item replenishing in cognitive diagnostic computerized adaptive testing. *Acta Psychologica Sinica*, *43*(7), 836–850.

[陈平, 辛涛. (2011b). 认知诊断计算机化自适应测验中的项目增补. *心理学报*, *43*(7), 836–850.]

Chen, P., Xin, T., Wang, C., & Chang, H. (2012). Online calibration methods for the DINA model with independent attributes in CD–CAT. *Psychometrika*, *77*(2), 201–222.

Chen, Y., Liu, J., & Ying, Z. (2015). Online item calibration for Q–matrix in CD–CAT. *Applied Psychological Measurement*, *39*(1), 5–15.

Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD–CAT. *Psychometrika*, *74*(4), 619–632.

Chiu, C.–Y. (2013). Statistical refinement of the Q–matrix in cognitive diagnosis.

Applied Psychological Measurement, *37*(8), 598–618.

de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, *76*(2), 179–199.

de la Torre, J., & Chiu, C. Y. (2016). General method of empirical Q–matrix validation. *Psychometrika*, *81*(2), 253–273.

de la Torre, J., & Lee, Y. S. (2010). A note on the invariance of the DINA model parameters. *Journal of Educational Measurement*, *47*(1), 115–127.

de la Torre, J., van der Ark, L. A., & Rossi, G. (2018). Analysis of clinical data from a cognitive diagnosis modeling framework. *Measurement and Evaluation in Counseling and Development*, *51*(4), 281–296.

Fan, J., & Li, R. (2001). Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American statistical Association*, *96*(456), 1348–1360.

Fan, J., & Lv, J. (2010). A selective overview of variable selection in high dimensional feature space. *Statistica Sinica*, *20*(1), 101–148.

Fan, Y., & Tang, C. Y. (2013). Tuning parameter selection in high dimensional penalized likelihood. *Journal of the Royal Statistical Society: SERIES B: Statistical Methodology*, *75*(3), 531–552.

Henson, R., Templin, J., & Willse, J. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, *74*(2), 191–210.

Hou, L. (2013). *Differential item functioning assessment in cognitive diagnostic modeling* (Unpublished doctoral dissertation). University of Delaware.

Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, *25*(3), 258–272.

Kang, H. A., Zheng, Y., & Chang, H. H. (2020). Online calibration of a joint model of item responses and response times in computerized adaptive testing. *Journal of Educational and Behavioral Statistics*, *45*(2), 175–208.

Klein Entink, R. H., Kuhn, J.–T., Hornke, L. F., & Fox, J.–P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times.

Psychological Methods, *14*(1), 54–75.

Li, H. (2012). *Statistical learning method*. Beijing: Tsinghua University Press.
[李航. (2012). *统计学习方法*. 北京: 清华大学出版.]

Lin, C. J., & Chang, H. H. (2019). Item selection criteria with practical constraints in cognitive diagnostic computerized adaptive testing. *Educational and Psychological Measurement*, *79*(2), 335–357.

Liu, H., You, X., Wang, W., Ding, S., & Chang, H. (2013). The development of computerized adaptive testing with cognitive diagnosis for an English achievement test in China. *Journal of Classification*, *30*(2), 152–172.

Ma W., & de la Torre, J. (2020). "GDINA: An R package for cognitive diagnosis modeling." *Journal of Statistical Software*, *93*(14), 1–26.

Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, *83*, 963–990.

Peng, S., Wang, D., Gao, X., Cai, Y., & Tu, D. (2019). The CDA–BPD: retrofitting a traditional borderline personality questionnaire under the cognitive diagnosis model framework. *Journal of Pacific Rim Psychology*, *13*, 1–14.

Rupp, A. A., & Templin, J. L. (2008). The effects of Q–matrix misspecification on parameter estimates and classification accuracy in the DINA model. E*ducational and Psychological Measurement*, *68*(1), 78–96.

Shi, S. S. (2017). *Cognitive diagnosis of Internet addition and its CD–CAT study* (Unpublished master's thesis). Jiangxi Normal University, Nanchang, China
[史双双. (2017). *网络成瘾的认知诊断及其 CD–CAT 的研究* (硕士学位论文). 江西师范大学, 南昌.]

Stocking, M. L. (1988). Scale drift in on–line calibration. *ETS Research Report Series, 1988*(1), 1–122.

Tan, Q., Cai, Y., Luo, F., & Tu, D. (2022). Development of a high–accuracy and effective online calibration method in CD–CAT based on gini index. *Journal of Educational and Behavioral Statistics*, *48*(1), 103–141.

Tan, Q., Wang, D., Luo, F., Cai, Y., & Tu, D. (2021). A high–efficiency and new online calibration method in CD–CAT based on information gain of entropy and

EM algorithm. *Acta Psychologica Sinica*, *53*(11), 1286–1300.

[谭青蓉, 汪大勋, 罗芬, 蔡艳, 涂冬波. (2021). 一种高效的 CD–CAT 在线标定新方法：基于熵的信息增益与 EM 视角. *心理学报*, *53*(11), 1286–1300.]

Tan, Z., de La Torre, J., Ma, W., Huh, D., Larimer, M. E., & Mun, E.–Y. (2023). A tutorial on cognitive diagnosis modeling for characterizing mental health symptom profiles using existing item responses. *Prevention Science: The Official Journal of the Society for Prevention Research*, *24*(3), 480–492.

Tang, F., & Zhan, P. (2021). Does diagnostic feedback promote learning? Evidence from a longitudinal cognitive diagnostic assessment. *AERA Open*, *7*(3), 296–307.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11*(3), 287–305.

Tu, D., Gao, X., Wang, D., & Cai, Y. (2017). A new measurement of Internet addiction using diagnostic classification models. *Frontiers in Psychology*, *8*, 1768.

van der Linden, W. J., Klein Entink, R. H., & Fox, J.–P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement*, *34*(5), 327–347.

Wainer, H., & Mislevy, R. J. (1990). Item response theory, item calibration, and proficiency estimation. In H. Wainer (Ed.), *Computerized adaptive testing*: *A primer* (Chap. 4, pp. 65–102). Hillsdale, NJ: Erlbaum.

Wang, D., Gao, X., Cai, Y., & Dongbo, T. U. (2020). A method of Q–matrix validation for polytomous response cognitive diagnosis model based on relative fit statistics. *Acta Psychologica Sinica, 52*(1), 93–106.

[汪大勋, 高旭亮, 蔡艳, 涂冬波. (2020). 基于类别水平的多级计分认知诊断 Q 矩阵修正：相对拟合统计量视角. *心理学报*, *52*(1), 93–106.]

Wang, H., Li, R., & Tsai, C. L. (2007). Tuning parameter selectors for the smoothly clipped absolute deviation method. *Biometrika*, *94*(3), 553–568.

Wang, W., Song, L., Ding, S., Meng, Y., Cao, C., & Jie, Y. (2018). An EM–based method for Q–matrix validation. *Applied Psychological Measurement*, *42*(6), 446–459.

Xi, C., Cai, Y., Peng, S., Lian, J., & Tu, D. (2020). A diagnostic classification version of Schizotypal Personality Questionnaire using diagnostic classification models. *International Journal of Methods in Psychiatric Research*, *29*(1), 1–16.

Xu, G., Wang, C., & Shang, Z. (2016). On initial item selection in cognitive diagnostic computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, *69*(3), 291–315.

Zhang X. G. (2010). *Pattern Recognitive (Third Edition)*. Tsinghua University Press.

[张学工. (2010). *模式识别(第三版)*. 清华大学出版社.]

Zhang, Y., Li, R., & Tsai, C. L. (2010). Regularization parameter selections via generalized information criterion. *Journal of the American Statistical Association*, *105*(489), 312–323.

Zheng, C., & Chang, H. H. (2016). High–efficiency response distribution–based item selection algorithms for short–length cognitive diagnostic computerized adaptive testing. *Applied Psychological Measurement*, *40*(8), 608–624.

Zou, H., & Li, R. (2008). One–step sparse estimates in nonconcave penalized likelihood models. *Annals of Statistics, 36*(4), 1509–1533.

Development of Online Calibration Method Based on SCAD Penalty and EM Perspective in CD-CAT: a study based on the G-DINA model

TAN Qingrong[1], CAI Yan[1], WANG Daxun[1], LUO Fen[1], TU Dongbo[1]
([1]*Department of Basic Psychology, College of Psychology, Army Medical University, Chongqing 400000, China*)
([2]*School of Psychology, Jiangxi Normal University, Nanchang 330022, China*)
([3]*College of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China*)

Abstract

Cognitive diagnostic computerized adaptive testing (CD-CAT) provides a detailed diagnosis of an examinee's strengths and weaknesses in the content measured in a timely and accurate manner, which can be used as a reference for further study or remediation planning, thus meeting the practical need for efficient and detailed test results. The successful implementation of CD-CAT is based on an item bank, but its maintenance is a very challenging task. A psychometrically popular choice for maintaining an item bank is online calibration. Currently, the research on online calibration methods in the CD-CAT that can calibrate Q-matrix and item parameters simultaneously is very weak. The existing methods are basically developed based on the deterministic input, noisy and gate (DINA) model. Compared with the DINA model, the generalized DINA (G-DINA) model has been more widely applied because it is less restrictive and can meet the requirements of a large number of test data in psychological and educational assessment. Therefore, if the online calibration method that jointly calibrates the Q-matrix and item parameters can be developed for models with few constraints such as G-DINA, its meaning is understood without explanation.

In current study, a new online calibration method, SCADOCM, was proposed, which was suitable for the G-DINA model. The construction of SCADOCM was based on the smoothly clipped absolute deviation penalty (SCAD) and marginalized maximum likelihood estimation (MMLE/EM) algorithm. For the new item *j*, the log-likelihood function with SCAD can be formulated based on the examinees' responses in this item and the examinees' attribute marginal mastery probability, and the q-vector of the new item can be estimated by the q-vector estimator based on SCAD. Then, the EM algorithm was used to estimate the item parameter of the new item

j based on the posterior distributions of examinees' attribute patterns, the examinees' responses to new item *j* and the estimated q-vector.

To examine the performance of the proposed SCADOCM and compare it with the SIE method, two simulation studies (Study 1 and Study 2) are conducted. Study 1 is based on a simulated item bank while Study 2 is based on the real item bank (Internet addiction item bank; Shi, 2017). In these simulation studies, four factors were manipulated: the calibration sample size (n_j = 50 vs. 100 vs. 500 vs. 1000 vs. 2000), the distribution of the attribute pattern (uniform distribution vs. high-order distribution vs. normal distribution), the item quality (U (0.05, 0.15) vs. U (0.1, 0.3)), and the online calibration methods (SCADOCM vs. SIE). The results showed that (1) SCADOCM has satisfactory calibration accuracy and calibration efficiency, and is superior to the SIE method. In addition, the traditional SIE method is not applicable for the G-DINA model, and its Q-matrix estimation accuracy rate is low under all experimental conditions. (2) The item calibration accuracy of SCADOCM and SIE increases with the increase of calibration sample and item quality under most conditions, and its item calibration accuracy in the uniform distribution/higher-order distribution is greater than that in the normal distribution. (3) The calibration efficiency of SCADOCM decreases with the increase of calibration samples, but it is less affected by the item quality and the attribute pattern distribution; the calibration efficiency of SIE decreases with the increase of calibration samples, but it is less affected by the item quality. Moreover, the calibration efficiency of the SIE method in the normal distribution is slightly slower than that of uniform distribution/high-order distribution.

To sum up the results, this study demonstrated that the SCADOCM has higher item calibration accuracy and calibration efficiency, and outperforms the SIE method; meanwhile, the traditional SIE method is not suitable for G-DINA model. All in all, this study provides an efficient and accurate method for item calibration in CD-CAT, and provides important support for further promoting the application of CD-CAT in practice.

Key words: Cognitive Diagnostic Computerized Adaptive Testing, Online Calibration, Q-matrix, G-DINA model, SCAD Penalty

附录 网络成瘾题库项目参数值

网络成瘾题库项目参数值如附表 1 所示，其中 P(0)、P(1)、P(00)、P(10)、P(01)、P(11)、P(000)、P(100)、P(010)、P(001)、P(110)、P(101) 、P(011)和 P(111) 表示缩减属性掌握模式(若题目测量 9 个属性中的前 2 个属性 $\boldsymbol{q}_j=(1,1,0,0,0,0,0,0,0)$，缩减属性掌握模式为 $\boldsymbol{\alpha}_{cj}^{*}=((0,0),(1,0),(0,1),(1,1))^{T}$)下被试的正确作答概率。如，$P$(0)和 P(1)分别表示题目测量 9 个属性中的某 1 个属性时，未掌握该属性的被试的正确作答概率和掌握该属性的被试的正确作答概率；P(10)表示题目测量 9 个属性中的某 2 个属性时，掌握 2 个属性中的第 1 个属性但未掌握第 2 个属性的被试的正确作答概率；P(011)表示题目测量 9 个属性中的某 3 个属性时，掌握 3 个属性中的第 2 个和第 3 个属性但未掌握第 1 个属性的被试的正确作答概率。

附表 1 题库项目参数值

题号	P(0)	P(1)	题号	P(0)	P(1)	题号	P(0)	P(1)
1	0.298	0.563	74	0.191	0.745	147	0.096	0.711
2	0.132	0.5	75	0.019	0.5	148	0.062	0.675
3	0.072	0.5	76	0.072	0.569	149	0.009	0.5
4	0.046	0.5	77	0.011	0.5	150	0.021	0.606
5	0.185	0.54	78	0.014	0.5	151	0.071	0.589
6	0.125	0.558	79	0.045	0.5	152	0.038	0.615
7	0.175	0.544	80	0.019	0.5	153	0.048	0.532
8	0.174	0.62	81	0.022	0.5	154	0.01	0.5
9	0.164	0.642	82	0.006	0.5	155	0.025	0.5
10	0.073	0.5	83	0.011	0.5	156	0.04	0.5
11	0.49	0.763	84	0.07	0.629	157	0.009	0.61
12	0.208	0.653	85	0.037	0.5	158	0.024	0.5
13	0.026	0.5	86	0.014	0.5	159	0.041	0.533
14	0.131	0.563	87	0.014	0.5	160	0.036	0.5
15	0.433	0.772	88	0.05	0.631	161	0.04	0.5
16	0.116	0.613	89	0.023	0.5	162	0.028	0.5
17	0.318	0.635	90	0.01	0.5	163	0.024	0.544
18	0.024	0.5	91	0.069	0.519	164	0.005	0.5
19	0.017	0.5	92	0.038	0.5	165	0.055	0.5
20	0.069	0.5	93	0.057	0.516	166	0.032	0.5
21	0.068	0.5	94	0.192	0.661	167	0.037	0.507

22	0.107	0.5	95	0.185	0.639	168	0.008	0.5
23	0.17	0.682	96	0.109	0.504	169	0.014	0.574
24	0.125	0.668	97	0.005	0.5	170	0.139	0.661
25	0.173	0.627	98	0.051	0.514	171	0.082	0.592
26	0.038	0.5	99	0.063	0.577	172	0.012	0.579
27	0.193	0.67	100	0.037	0.5	173	0.02	0.553
28	0.054	0.5	101	0.145	0.5	174	0.042	0.648
29	0.102	0.5	102	0.067	0.618	175	0.012	0.592
30	0.209	0.5	103	0.128	0.53	176	0.017	0.5
31	0.383	0.5	104	0.061	0.558	177	0.039	0.5
32	0.092	0.5	105	0.021	0.5	178	0.042	0.5
33	0.029	0.5	106	0.026	0.5	179	0.035	0.53
34	0.032	0.5	107	0.152	0.5	180	0.042	0.517
35	0.277	0.773	108	0.014	0.5	181	0.012	0.519
36	0.127	0.5	109	0.095	0.66	182	0.008	0.5
37	0.123	0.5	110	0.028	0.5	183	0.014	0.5
38	0.061	0.536	111	0.033	0.549	184	0.013	0.568
39	0.05	0.5	112	0.127	0.643	185	0.06	0.5
40	0.15	0.592	113	0.073	0.5	186	0.073	0.574
41	0.032	0.5	114	0.018	0.5	187	0.072	0.613
42	0.27	0.839	115	0.028	0.5	188	0.026	0.5
43	0.062	0.5	116	0.022	0.5	189	0.018	0.5
44	0.237	0.74	117	0.007	0.5	190	0.028	0.5
45	0.063	0.5	118	0.04	0.5	191	0.014	0.5
46	0.094	0.5	119	0.114	0.5	192	0.015	0.5
47	0.117	0.623	120	0.05	0.579	193	0.041	0.5
48	0.041	0.5	121	0.012	0.5	194	0.009	0.5
49	0.262	0.697	122	0.043	0.632	195	0.033	0.5
50	0.042	0.5	123	0.025	0.514	196	0.099	0.659
51	0.064	0.522	124	0.051	0.5	197	0.013	0.5
52	0.011	0.5	125	0.019	0.5	198	0.023	0.5
53	0.028	0.5	126	0.035	0.551	199	0.026	0.5
54	0.009	0.5	127	0.032	0.5	200	0.005	0.532
55	0.067	0.568	128	0.079	0.723	201	0.011	0.5
56	0.026	0.5	129	0.083	0.674	202	0.039	0.52
57	0.026	0.5	130	0.052	0.575	203	0.071	0.524
58	0.056	0.5	131	0.027	0.627	204	0.051	0.5
59	0.094	0.68	132	0.025	0.5	205	0.21	0.64
60	0.175	0.809	133	0.211	0.617	206	0.009	0.5
61	0.099	0.512	134	0.051	0.5	207	0.056	0.59
62	0.046	0.549	135	0.019	0.5	208	0.045	0.5
63	0.077	0.72	136	0.02	0.5	209	0.034	0.5
64	0.124	0.607	137	0.011	0.5	210	0.018	0.5

65	0.066	0.5	138	0.01	0.5	211	0.037	0.5
66	0.061	0.5	139	0.087	0.677	212	0.135	0.612
67	0.012	0.5	140	0.038	0.616	213	0.025	0.5
68	0.189	0.697	141	0.02	0.5	214	0.018	0.5
69	0.026	0.5	142	0.059	0.635	215	0.018	0.5
70	0.172	0.679	143	0.007	0.5	216	0.058	0.5
71	0.046	0.621	144	0.019	0.5	217	0.029	0.5
72	0.012	0.5	145	0.029	0.603			
73	0.5	0.76	146	0.064	0.628			

题号	*P*(00)	*P*(10)	*P*(01)	*P*(11)
218	0.289	0.435	0.312	0.525
219	0.161	0.4	0.325	0.575
220	0.151	0.576	0.403	0.611
221	0.134	0.445	0.367	0.729
222	0.055	0.255	0.078	0.508
223	0.036	0.187	0.235	0.5
224	0.013	0.118	0.039	0.5
225	0.181	0.51	0.569	0.644
226	0.067	0.344	0.382	0.561
227	0.048	0.143	0.271	0.5
228	0.055	0.24	0.272	0.5
229	0.011	0.093	0.186	0.504
230	0.088	0.202	0.491	0.649
231	0.045	0.104	0.564	0.583
232	0.013	0.093	0	0.5
233	0.022	0.178	0.057	0.5
234	0.007	0.028	0.037	0.5
235	0.034	0.168	0.396	0.536
236	0.019	0.13	0.158	0.588
237	0.012	0.072	0	0.5
238	0.015	0.054	0.154	0.52
239	0.032	0.319	0.084	0.5
240	0.219	0.494	0.4	0.5
241	0.006	0.021	0.055	0.5
242	0.008	0.074	0.102	0.5
243	0.01	0.072	0.108	0.5
244	0.015	0.196	0.059	0.612
245	0.06	0.191	0.341	0.652
246	0.004	0.128	0.039	0.5
247	0.007	0.167	0.174	0.546
248	0.005	0.039	0.229	0.568
249	0.009	0.216	0.084	0.533

250	0.016	0.114	0.335	0.621
251	0.008	0.151	0.021	0.5
252	0.011	0.137	0.072	0.5
253	0.012	0.08	0.125	0.5
254	0.42	0.428	0.414	0.5
255	0.038	0.131	0.372	0.576
256	0.039	0.39	0.213	0.611
257	0.016	0.076	0.229	0.546
258	0.016	0.069	0.233	0.542

题号	$P(000)$	$P(100)$	$P(010)$	$P(001)$	$P(110)$	$P(101)$	$P(011)$	$P(111)$
259	0.116	0.602	0.325	0.326	0.482	0.345	0.34	0.62
260	0.01	0.072	0.11	0	0.167	0.12	0.121	0.5
261	0.036	0.285	0.252	0.575	0.376	0.38	0.322	0.639
262	0.031	0.254	0.17	0.515	0.148	0.262	0.486	0.604
263	0.009	0	0.019	0.037	0.065	0.162	0.147	0.5